

# How to Map Shopper Preferences to Purchase Intent

## Getting to Win-Win

With the majority of eCommerce brands focusing on personalization and customer journey optimization, it's valuable to unravel the underlying methodology that determines predicted affinities and intents of our shoppers.

The entire purpose of the personalization strategy is to understand shoppers, thereby allowing you to provide them the most relevant experience, and in turn, to help you build your business. Ideally it's a win-win, for you and your customers.

World-class investor and Warren Buffet's partner, Charlie Munger inspires us to achieve this mutual-benefit business ideal:



***“What really works in life is win-win and that requires some sensitivity to the other fellows' way of thinking and his needs. ...It really starts working when two people trust each other.”***

So to take a page out of Munger's book and provide a relevant, personalized, “win-win” experience, we start by building an understanding of our customers.

# An Involved Model

The four pillars of developing and launching shopper-personalized journeys are:

Strategy → Analysis → Decision → Action

In this guide, we dig into the **Analysis** category. We'll gather a better comprehension of the methodology and the data types employed when designing and building these intent-based customized shopper journeys.

---

When building a model for scoring lead quality, often we break down the scoring elements into two simple buckets.

- **Behavioral** → *What they actually do*
- **Demographic / Context** → *Everything else about them and their environment*

However, with personalization in a digital world – one filled with massive troves of data, fickle preferences, high advertising costs, and evolving software – the model breakdown will necessarily be a bit more involved.

In the following pages we'll break down the types of data and systematic approach to predictive modeling in eCommerce.



This modeling determines the intent, affinities, and purchase path of customers: to provide better experiences for shoppers and efficient conversion optimization for your store.



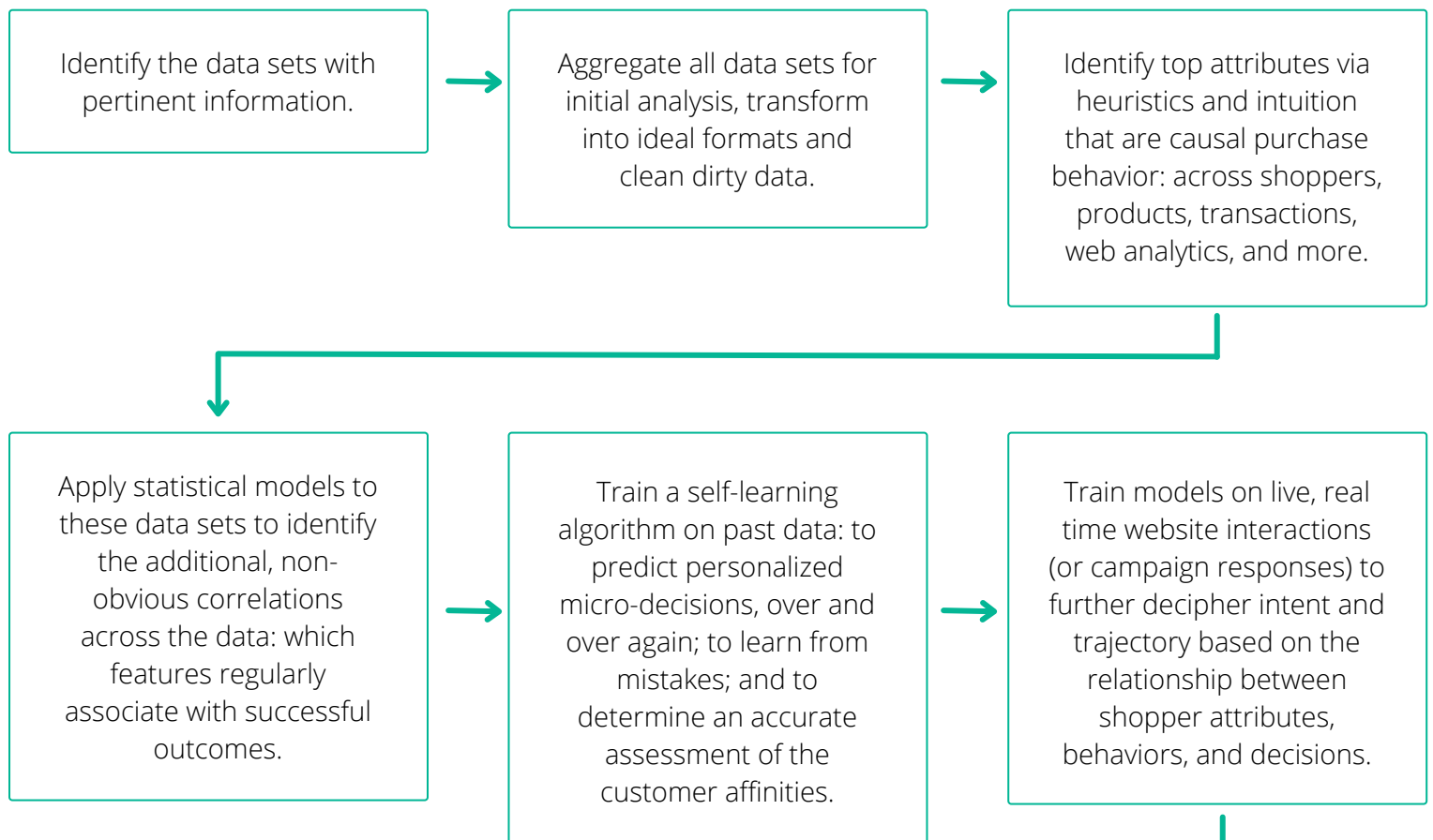
# Methodology

The systematic approach used for this particular problem set involves creating microsegmentation of your customer audiences and determining the statistical correlations between many input variables and the many possible outcomes.

The process is a mix of: data + statistics + automated testing + recursive iteration.

We pull as much relevant data as possible, determine those correlations, test the accuracy and efficacy, and then iterate over and over.

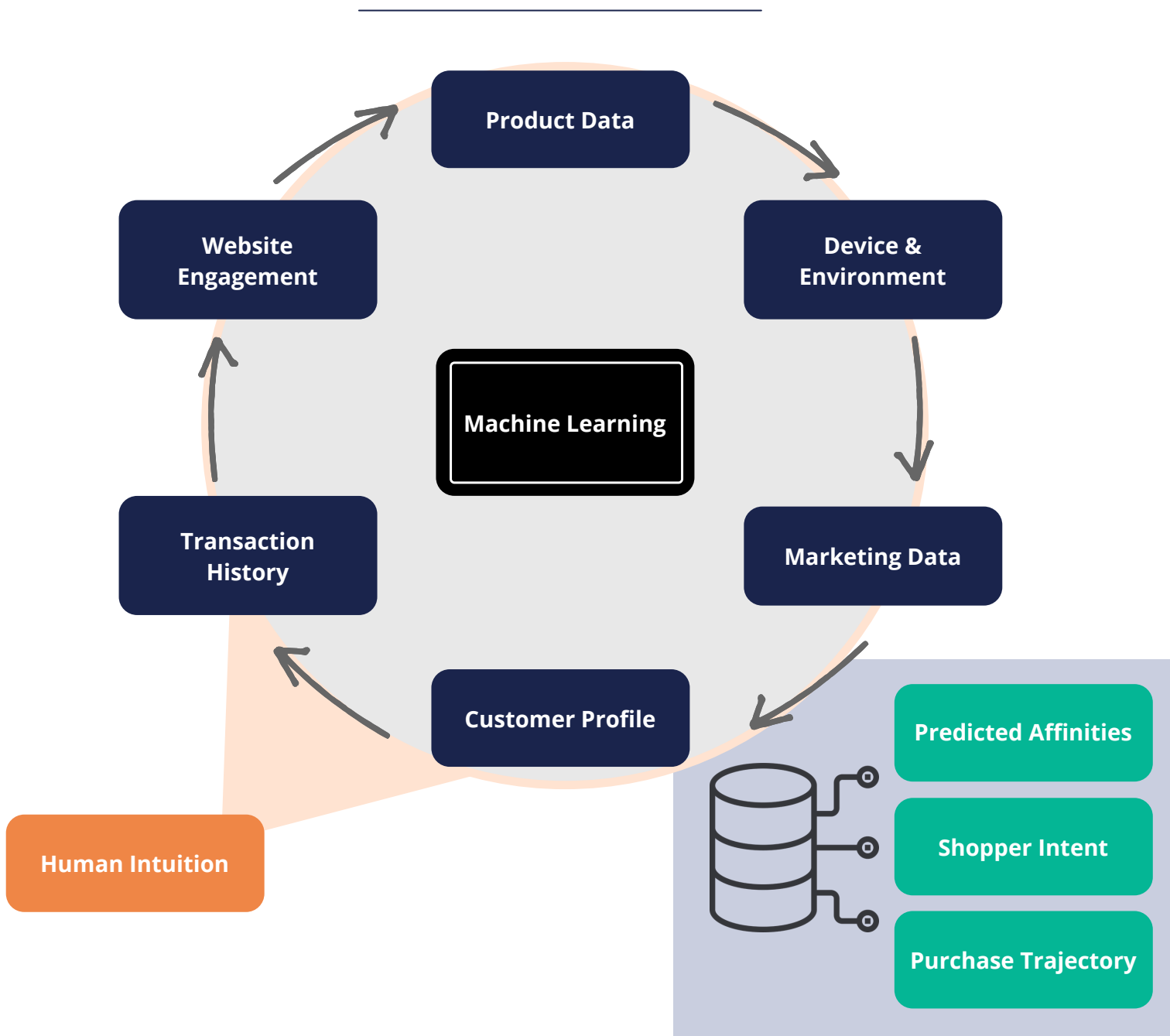
**Let's spell this out in a bit more detail.**



Through this methodology you're able to understand your shoppers in a highly granular way – and, of course, in a way that allows you to activate those insights for your business.

At the risk of overstating the obvious: more data is extremely valuable for analysis. It takes us closer and closer to “reality”, which we are aiming to predict.

Thus, you'll notice a wide array of data types that fit into the methodology. Let's dig in.



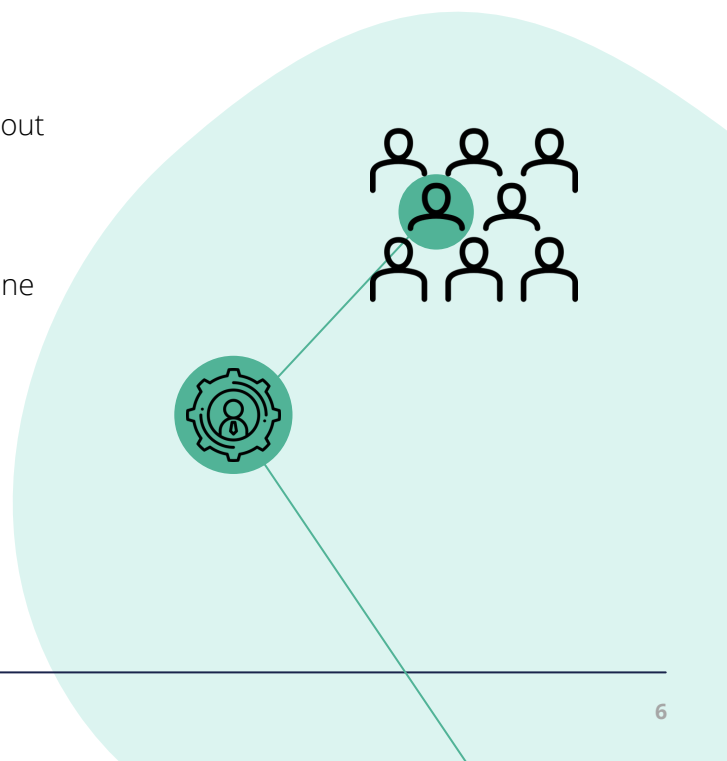
Data Type:

# Customer Profile

- **Purchase History**
- **Add to Cart Patterns**
- **Survey Responses**
- **Product Preferences**
- **CRM Attributes**
- **Shopping Journey**
- **Purchase Frequency**

The customer profile looks at the purchase history, product preferences, and any other data collected about a known customer. Many sellers' CRMs have helpful contextual and demographic info saved.

These folks have purchased in the past and provide one of the most powerful datasets to connect the dots between other variables and purchase behavior.



Data Type:

# Transaction History

- Order Values
- Chargebacks & Refunds
- Payment Types
- Shipping / Billing Geography
- Transaction Metadata



Transactions are separated out from customer profiles and their purchase history because this data set alone has very interesting attributes.

Not only does this include the refunds, returns, and that impact on dollars, but it also includes the metadata around transactions – **they aren't necessarily obvious, but are potentially powerful.**

An example: the credit card provider or payment method on a transaction may correlate strongly with certain shopping behaviors or product affordability – seemingly irrelevant but potentially actionable.

Data Type:

# Website Engagement

- **Browsing Patterns**
- **Traffic Source & Channel Detail**
- **Promotion Engagement**
- **Content Viewed**
- **Time on Page & Scroll Depth**
- **Abandoned Cart Analysis**
- **Similar Visitors**

This is one of the most central data sets and can be used – standalone – to power significant personalization. All behavior and engagement with your web store is tracked... and is extremely informative and actionable.

Looking into the browsing patterns and how they relate to 'Add to Cart' actions versus 'Abandoned Cart' – or how the traffic source relates to content viewed – all provides powerful insight to use for real-time personalization.

**Notably, the similarities between current website visitors and past shoppers define a mapping between other “second-degree attributes”** (like Device & Environment categories described below) back to subscriptions or purchases.





Data Type:

# Product Data

- **Product Taxonomy**
- **SKU Details**
- **Product Prices & Discounts**
- **“Frequently Bought Together” Relevance**
- **Product Images**
- **Reviews and Popularity**

Your products are all categorized and detailed with tons of attributes. These enable you to **match product specifics to the shoppers who bought them** and to other products that have first-, second-, or third-degree similarities.

Image recognition can learn deeper relationships between the products than the hard data. The industry is still learning, but images may have larger than expected impacts on how users engage with that product, given the visual stimulation.



Prices and discounts are important here to align with each shopper’s purchasing power. Reviews and survey responses obviously provide explicit information about affinity and satisfaction.

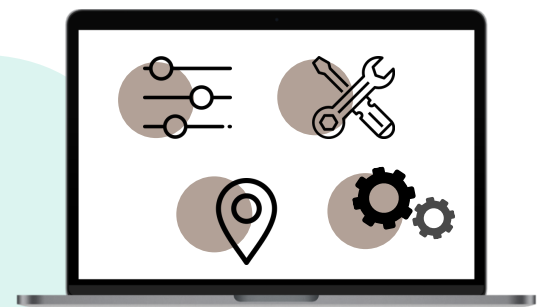
Data Type:

## Device & Environment

- **Device Type**
- **Code Running**
- **Geography & IP**
- **Browser Data**
- **Operating Systems**
- **Date/Time Patterns**

The device and technical environment that a customer uses to engage with your brand has deep implications. Geography is a critical attribute... as is date/time. These more obvious data points around a shoppers visit, click, or purchase will inform pretty straightforward optimization strategies (think: geo-based targeting, and time-of-day bid adjustments, etc).

The other attributes, like which browser or operating system is used, can **unearth interesting relationships between site visitors and their browsing behavior or purchase likelihood**. These wouldn't be direct causal features, but are indicators that – when combined with other data sets – enrich the accuracy of predictions.



Data Type:

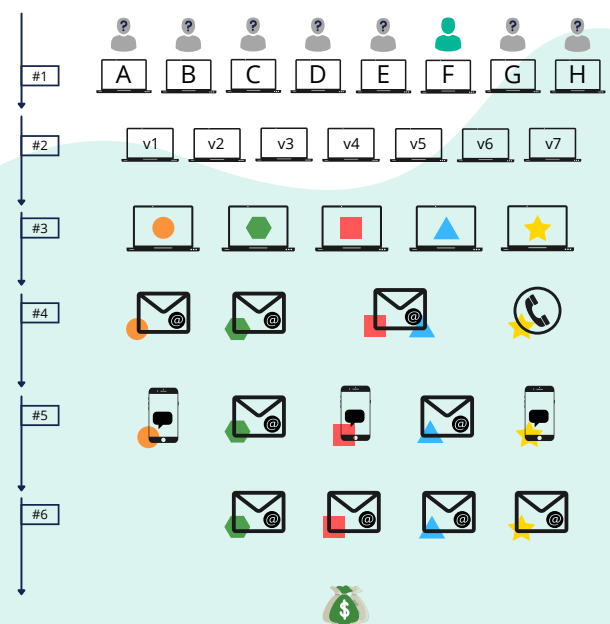
# Marketing Data

- Promotion History & Performance
- Campaign Engagement
- Paid Channel Analytics
- Direct Marketing Campaigns
- Conversion Data

The success and failure of marketing efforts are very helpful indicators to understand customer behavioral trends. For our current goals, we're interested in the relationship between campaign engagement rates and:

- A) the shopper profiles who **did or did not interact**, and then
- B) **how deeply** each person engaged.

Conversions for various email promotions, content messaging, creative visual design, or channels more broadly should be powerful lessons for marketers on affinities of their audiences.



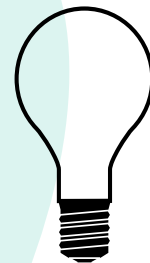
Input Type:

# Human Intuition

- **Heuristics**
- **Innate Relationships**
- **Pre-populated Weighting**

We'll give this a bit more attention shortly; however, the core is that some affinities and intentions are pretty intuitive – **even common sense.**

We can use heuristics (simple, generally true shortcuts) to kick-off modeling, and weigh those heavily up front.



## Applying human experience

You're tracking and capturing all of the required data: check. You've integrated the systems into a central tool: check.

At this point we might think: let's just let the machine learning start analyzing, labeling, modeling, and training.

**However... before software intelligence is applied, we apply human intelligence.**

Why? There are relationships that humans know innately, via intuition, or through life experience that can basically be pre-loaded into the algorithm.

For us (organic, carbon-based lifeforms) many of these relationships can be obvious, but they will fast-track the machine's ability to learn and get to better results more quickly with less testing.

**It would be foolish to ignore correlations that intuitively have a high coincidence rate.** Some easy examples are gender, sizes, inventory levels, seasonality impact.

Don't underestimate the impact of introducing these relationships that are "simple and easy" for humans.

Software views the problem of optimization mathematically, and thus will need to *prove* the simple correlation with statistics and data: you can give the machine a head start and it can then tweak the weighting as appropriate from there.

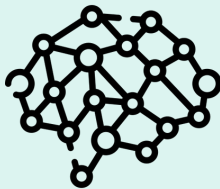
Using heuristics as a starting point to predict shopper affinities, preferences, and probability to purchase gets us a step closer – and it's worth it.

# Applying machine learning

**Machine learning can be an entire book by itself, but the short and sweet steps involved in the methodology for determining customer intent and affinities are as follows:**

- Create infrastructure for data plumbing, task orchestration, and feedback loops
- Define the architecture of the learning algorithm, applying statistical methodologies
- The machine computes the correlating features (the variable attributes of each data set that may have some impact on the predicted outcome), and the target labels (the outcomes)
- Train the machine: run the models on actual past data – like revenue, purchases, conversions – to predict outcomes based on inputs (or, in short: predict labels based on features). Depending on the degree of accuracy or inaccuracy, the model will adjust to improve, over and over.
- Go “live” and run this on real, live data moving forward – the machine will continue to train.

This is the “sausage making” that can be a bit esoteric, so we’ll leave it at that for this guide.



$$\mathbf{A} \cdot \mathbf{B} = \|\mathbf{A}\| \|\mathbf{B}\| \cos \theta$$

$$\text{cosine similarity} = S_C(A, B) := \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}}$$

$$A = [A_1, A_2]^T, \text{ then } \bar{A} = \left[ \frac{(A_1 + A_2)}{2}, \frac{(A_1 + A_2)}{2} \right]^T, \text{ so } A - \bar{A} = \left[ \frac{(A_1 - A_2)}{2}, \frac{(-A_1 + A_2)}{2} \right]^T$$

\*These stats formulas are incomplete. Source: Wikipedia

## Tying it together

So how do you analyze all of your various data types to learn about your customer's preferences and purchase pathways? Now you know.

The next question might be: how do you use that unearthed info to make decisions that lead to business results? A topic for another guide! (Or a [conversation](#) with the Jarvis ML team!)

The methodology and data sets we've walked through in this guide should paint the picture of how we as marketers, analysts, and eCommerce managers should go about the process of understanding shopper affinities and intent.

In reality, the lionshare of this operation would be done by software: intelligence engines and end-to-end personalization platforms.

### Summarizing the process we learned:

By utilizing a wide array of data, **we can correlate attributes (behavioral, contextual, etc) with conversion and purchase expectations** for leads and customers. We can then use these predicted relationships to determine paths of intent and directions of affinity for each known and brand new shopper.

The more data and more time we apply to this methodology, the better the accuracy of the predictions. And that leads to shoppers experiencing the most relevant content, and your business improving conversions and revenue. Personalization is a tool for the savvy – and you're now armed with knowledge of how to explore it.



**Jarvis ML guides companies on how to best utilize their business data to grow revenue and solve their most important challenges.**

Our machine learning intelligence engine organizes all of your transaction, product, customer, and marketing data to draw meaningful connections and power on-the-spot decision making. You'll be equipped to personalize your customer journey; calculate the highest-converting prices and products; and grow revenue predictably.

[Learn More about Jarvis ML](#)

[Request a Demo](#)