

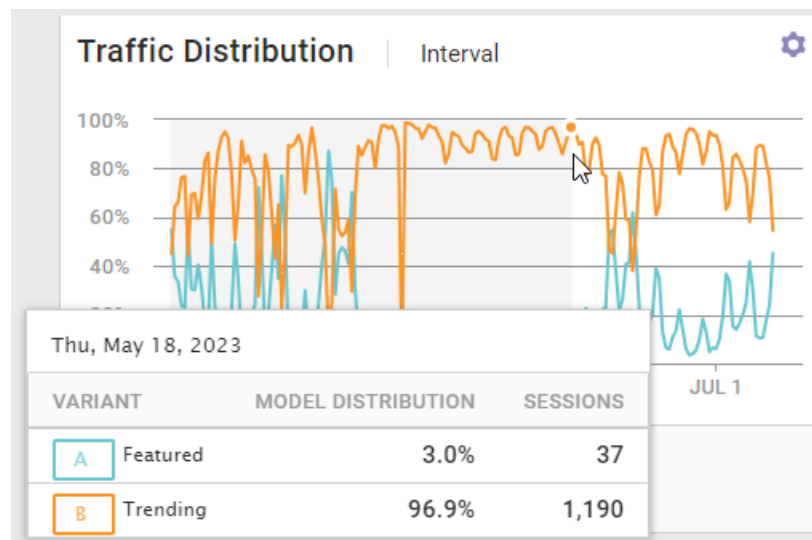
Understanding Automated Personalization Experience Analytics

The Automated Personalization experience shows the right content to each customer on your site. It is most effective when you have the time to allow an Automated Personalization experience to complete a state of learning. To get the most from an Automated Personalization experience, you must understand some background and tools to analyze the metrics in Monetate.

Traffic Allocation and Automated Personalization

In a standard A/B test, traffic almost always is distributed based on allocated fixed percentages. In the optimal case, the distribution is an even 50/50. In more conservative cases, the split might be 80/20.

With Automated Personalization experiences, the algorithm optimizes toward a goal metric and does not consider customer context.



During the exploration phase, when the Personalization engine actively investigates the best variant for your site visitors at the start of the experience, you will most likely see even traffic allocation. Even traffic allocation could occur in the exploitation phase, but it's generally less likely.

Lift and Chance Automated Personalization Is Better

The Automated Personalization experience has two primary performance metrics:

- **Lift** is the difference in the average cumulative performance of the holdout from the average cumulative performance of the Automated Personalization portion of the experience. When you observe that lift is positive for an Automated Personalization experience, you're actively personalizing.
- **Chance 1:1 Assignment is better than Random** is the probability that Automated Personalization is better than the holdout. The higher the lift, the greater the separation between holdout and Automated

Personalization and the higher the chance that Automated Personalization is better. It's possible to have a small lift with a large number of samples, which leads to greater confidence in the separation. This metric accounts for both variance and sample size.

Homepage Product Recs

Performance Engine Insights Real-Time Counts Recommendations

Start Date: January 19, 2023

★ Goal Metric: HP Product Rec Click

Run Time: 5 months 25 days 1 hour

Has the Engine learned enough to improve HP Product Rec Click?


NO MAYBE **YES** Currently, there's a 99.9% Chance Intelligent 1:1 Assignment is better than Random.

↑5.79% ^{+/-} 2.93% 1:1 ASSIGNMENT LIFT OVER RANDOM

+6 GOAL EVENTS / DAY MORE THAN RANDOM

+1,036 TOTAL GOAL EVENTS MORE THAN RANDOM

What should I do next

VARIANTS (2)	VARIANT ASSIGNMENT	HP PRODUCT REC CLICK	SESSIONS	RATE	CURRENT CHANCE IS BETTER
	80% Intelligent 1:1 Assignment	18,927	260,101	7.28%	99.9%
	20% Random Assignment	4,476	65,071	6.88%	0.1%

To better understand how lift and the chance that Automated Personalization is better work together, consider this scenario.

A lookout stands atop a castle wall and looks across a desert. His job is to spot potential threats as far in the distance as possible to give the castle's occupants as much forewarning as possible. However, the desert heat can cause the horizon often to appear hazy and wavy, making his ability to see what's incoming at times more difficult.

This uncertainty is similar to the concept of variance. The truth is affected by fluctuations in the data that make it difficult to understand reality from noise.

Consider too that the lookout can more easily identify any incoming threats from a greater distance if they are tall and ride horses. This factor is comparable to the concept of lift: If there's a greater distance between the lookout and the Automated Personalization experience, then the lookout can be more confident that he's observing an attacker instead of nothing at all and the higher the chance that Automated Personalization is better.

If an approaching threat crawled on all fours, the lookout would still see it but at a shorter distance. This loss of advance warning is a cost but one that is necessary to distinguish approaching threats from benign actors. For personalization purposes this shorter distance comes at the cost of more samples. More samples allow you to detect a smaller true lift overcoming variance.

Tactics for Different States of Analytics

The scenarios below are possible outcomes that you may see in the experience results for an Automated Personalization experience. While the Personalization Engine is sophisticated, you should avoid classifying outcomes in black and white terms. The analysis of each of the following outcomes can help guide you on what to do next.

High Lift, High Chance Automated Personalization Is Better, Uneven Traffic Distribution

When facing results that show high lift and a high chance that one-to-one assignment is better than random assignment, but traffic allocation is uneven, bear in mind that the Engine fits variants to audiences to maximize the goal metric. Therefore, if you're satisfied with the lift, then let the experience run so you can reap the rewards.

If you have an internal marketing calendar that disqualifies the experience content from eligibility, consider that any new content must go through a period of learning and exploration. Keep the content running as long as possible.

Low Lift, Low Chance Automated Personalization Is Better, Even Traffic Distribution

If the experience results show low lift, a low chance that one-to-one assignment is better than random assignment, and the traffic distribution is even, then the Engine is likely still in a learning and exploration phase. The experience may not have enough sessions, or there may not be enough polarization amongst the variants to exploit.

If the experience has fewer than 30,000 sessions per variant or has fewer than the total of 30,000 sessions multiplied by the number of variants, then let the experience continue to run. Each suggested session count isn't a hard threshold, but using one of them may give you some assurance that you're not operating on a decision before the Engine has a chance to learn.

If you've met that threshold and continue to observe low lift, low certainty, and an even traffic distribution, then you may consider re-examining whether the content is discoverable and polarized. If you believe an area of opportunity exists, then consider creating a new Automated Personalization experience with new and even more polarized creatives.

Additionally, consider whether the Engine hasn't considered other data points as context that may identify groups of customers who might react differently to the creative. For example, if you have an image focused on bathtubs and another focused on outdoor furniture, then think about the factors that might separate visitors' likelihood to engage. People in urban environments might have less need for outdoor furniture than consumers in rural and suburban environments. Consider leveraging more specific geographic data.

High Lift, High Chance Automated Personalization Is Better, Even Traffic Distribution

If you find that the experience results show high lift, a high chance that one-to-one assignment is better than random assignment, and even traffic distribution, then the Engine found that the variants are best fit to audiences that make up an even distribution while maximizing the experience's goal metric. The audiences best for each variant are relatively similar in size, and this situation isn't uncommon.

Let the experience continue to run as long as you can to reap the rewards.

High Lift, Low Chance Automated Personalization Is Better, Even or Uneven Traffic Distribution

When the experience results show high lift, a low chance that one-to-one assignment is better than random assignment, and the traffic distribution is even or uneven, then the Engine has found that the variants are best fit for certain audiences and has captured lift. However, it needs more data and time to increase the lift certainty. This lift may occur by chance, and you want to be certain. Therefore, let the experience continue to run for the time being.

If the analytics remain unchanged after multiple weeks and you've exceeded the 30,000-session threshold for each variant, then reconsider whether the experience content is discoverable and polarized.

If you observe that traffic allocation has diverged heavily from even, then the model learned something. Therefore, you might consider letting the experience run for another few weeks.

Low Lift, High Chance Automated Personalization Is Better, Most Traffic Allocated to One Variant

When the Engine has allocated most traffic to a single variant and there's a high chance that one-to-one assignment is better than random assignment, then the majority of your customers are best served by that variant.

However, you have a few options if you're nevertheless uncomfortable with the amount of lift you're seeing and want to continue to invest in this location of the customer experience.

You may have an opportunity to introduce more content variants to differentiate amongst the majority and to drive more value—other variants are best served to a smaller audience. Consider revisiting the premise that you leveraged around polarized creatives, and ensure you're embracing that concept. You may be able to push it even further.

Low Lift, High Chance Automated Personalization Is Better, Even Traffic Distribution

You need to determine whether you believe you can build bolder, more polarized content or if you're OK with the incremental lift that you've gained right now.

Negative Lift, Low Chance Automated Personalization Is Better, Even Traffic Distribution

When you see negative lift, a low chance that one-to-one assignment is better than random assignment, and even traffic distribution, know that the Automated Personalization experience is still learning. Both the holdout and the Automated Personalization experience are essentially random allocations, which means both are subject to change and the possibility of negative lift by chance.

If the experience has reached the suggested threshold of 30,000 sessions per variant and you still see the same results, then you may consider re-evaluating your strategy to ensure you have polarized the experiences enough.

Negative Lift, Low Chance Automated Personalization Is Better, Most Traffic Allocated to One Variant

Ensure the actions in each variant are properly configured. If one variant includes actions that target the home page and another variant includes actions that target a product detail page, this situation can cause this kind of

allocation misalignment.

If the configuration is correct, two other explanations are possible:

- The Engine might have observed a trend in behavior and learned from it even though it was actually random. This result is unlucky. The Engine automatically recovers, but separating performance enough from random to demonstrate lift will take some time.
- Behavior could have just meaningfully changed from what the Engine learned in the past. Here, the Engine is still exploiting its knowledge but hasn't yet corrected based on new observations.

Ensure you have enough sessions, or determine whether you want to continue running the experience or build a new one with more polarized content alternatives.