

Audience Validation Framework

Overview

This document describes the audience validation approach used to measure and compare the quality of audiences built on the Akkio + Affinity Solutions (AFS) data platform. The framework provides a standardized, repeatable methodology for evaluating how well any audience — whether a known-shopper seed, a lookalike expansion, or a propensity-scored cohort — predicts real consumer purchase behavior during a future holdout period.

The goal is to provide **apples-to-apples comparisons** across audience construction methods, enabling data-driven decisions about which approach delivers the highest-quality audiences for activation.

Note on platform capabilities: The audiences delivered in this POC were built using Akkio's lookalike modeling and a deterministic SQL-based similarity scoring approach. Akkio's platform also includes a **dedicated propensity modeling engine** and a **built-in RFM feature engine** that were not utilized in this engagement. These capabilities — purpose-built for purchase propensity prediction with automated feature engineering, model training, and temporal holdout management — can further improve audience quality, particularly for niche brands where general-purpose similarity scoring has structural limitations. Integrating these tools into the workflow is a natural next step beyond this POC.

Key Concepts

What We Measure

Every audience is evaluated against the same set of metrics, calculated over a **holdout window** — a time period that was deliberately excluded when the audience was built. This ensures we are measuring true predictive power, not just historical behavior.

Metric	Formula	What It Tells You
Total Audience IDs	Count of distinct IDs in the audience	Audience reach / size
Active Matched IDs	Audience IDs with any transaction in the holdout period	Observable panel coverage within the audience
Brand Shoppers	Audience IDs with at least one brand transaction in holdout	How many audience members actually purchased
Brand Transactions	Total brand transactions from audience members in holdout	Volume of purchase activity
Brand Spend	Total dollar amount of brand transactions in holdout	Revenue impact
Shop Rate	Brand Shoppers / Active Matched IDs	% of active audience members who purchased the brand
Spend Rate	Brand Spend / Active Matched IDs	Average brand spend per active audience member
Avg Ticket	Brand Spend / Brand Transactions	Average purchase size
Avg Transactions per Shopper	Brand Transactions / Brand Shoppers	Purchase frequency among converters

Why Active Matched IDs is the denominator (not Total Audience IDs):

Not every individual in the panel will have observable transaction activity in any given month. Using Total Active Matched IDs — audience members who had at least one transaction of *any* kind during the holdout — ensures we measure audience quality against observable behavior, rather than penalizing for panel coverage gaps.

Seed vs. Lookalike — Why Both Matter

Audience Type	What It Measures	Expected Behavior
Seed (Known Shoppers)	Retention / repeat purchase — "Do existing buyers keep buying?"	Higher shop rate (selecting on known buyers)
Lookalike / Propensity	Acquisition / prediction — "Can we find NEW brand shoppers before they buy?"	Lower shop rate, but demonstrates true predictive

Audience Type	What It Measures	Expected Behavior
		lift

A **seed audience** of known brand shoppers validated in a holdout window measures **retention** — these individuals already purchased, so the shop rate will be naturally high. This is useful for understanding audience stickiness but does not demonstrate predictive modeling power.

A **lookalike audience** or **propensity-scored cohort** of individuals who have *not* been observed purchasing the brand measures **acquisition** — the ability to identify future brand shoppers before they convert. This is the more meaningful test of audience quality and the direct comparison between modeling approaches.

Validation Methodology

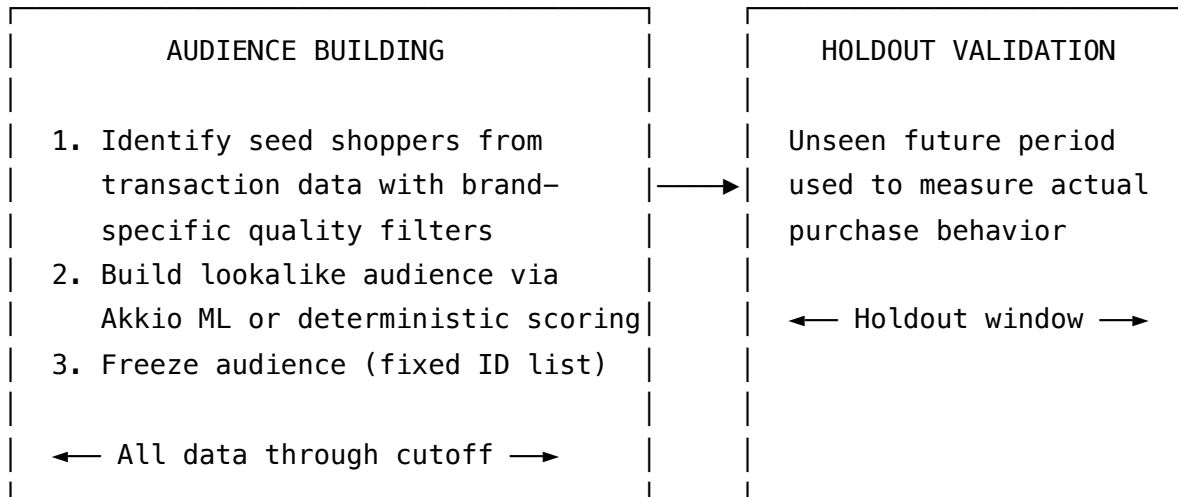
Timeline Split

The validation uses a temporal holdout design to measure audience quality against future purchase behavior:

Period	Role
All data through the build cutoff date	Used for audience construction (seed identification, feature engineering, model training)
Holdout window (post-cutoff)	Used exclusively for measuring audience performance — never seen during audience building

For the pre-delivery validation cycle, the holdout window is **September 1, 2025 through September 30, 2025**. AFS will conduct a definitive out-of-sample validation against the **October – December 2025** transaction window.

How Validation Works



Step-by-step process:

1. **Identify seed shoppers** — match brand keywords against transaction data with quality filters (minimum transaction thresholds, date bounds, holiday exclusions where applicable)
2. **Build the lookalike audience** — seed is used to generate a scored, ranked audience via Akkio's ML platform (LAL) or deterministic SQL similarity scoring across RFM + demographic + interest features
3. **Freeze the audience** — the list of IDs is fixed before looking at holdout data
4. **Join audience IDs** against all transactions in the holdout window
5. **Identify active members** — audience IDs that had *any* transaction (any brand) during the holdout
6. **Identify brand converters** — audience IDs that had at least one brand-specific transaction during the holdout
7. **Compute metrics** — Shop Rate, Spend Rate, Avg Ticket, Avg Transactions per Shopper, and Lift vs. general population baseline

Brand Matching Logic

Brand transactions are identified by matching keywords (case-insensitive) against three columns in the transaction data:

- BRAND_NAME
- STORE_NAME
- MERCHANT_DESCRIPTION

For example, to identify ActBlue transactions, the keyword `ACTBLUE` is matched against all three columns. Multiple keywords per brand are supported and OR'd together.

Delivered Audiences — Validation & Results

Overview

Fourteen final production audiences have been built, validated, and **delivered to the AFS S3 bucket** for activation. These audiences span four brands/segments — **ActBlue** (two construction methods), **Ross** (two construction methods), **BetMGM** (two construction methods), and **Holiday Discount Department Store Shoppers** — each with a compact (precision-optimized) and expanded (reach-optimized, 20M–30M) variant.

AFS will conduct their own out-of-sample validation against the **October – December 2025** transaction window, a period not available during audience construction.

Final Audience Inventory

#	Audience Name	Type	Size	Brand(s)
1	ActBlue Contributor Seed Lookalike - Final	Akkio LAL	874,141	ActBlue
2	ActBlue Contributor Seed Lookalike 20M - Final	Akkio LAL	20,161,266	ActBlue
3	ActBlue Deterministic Lookalike Audience - Final	Deterministic SQL Scoring	500,000	ActBlue
4	ActBlue Deterministic Lookalike Audience 20M - Final	Deterministic SQL Scoring	20,000,000	ActBlue
5	Ross Seed Audience Lookalike - Final	Akkio LAL	3,075,285	Ross

#	Audience Name	Type	Size	Brand(s)
6	Ross Seed Audience Lookalike 30M - Final	Akkio LAL	30,831,285	Ross
7	Ross Deterministic Lookalike Audience - Final	Deterministic SQL Scoring	500,000	Ross
8	Ross Deterministic Lookalike Audience 30M - Final	Deterministic SQL Scoring	30,000,000	Ross
9	BetMGM Seed Lookalike - Final	Akkio LAL	340,999	BetMGM
10	BetMGM Seed Lookalike 20M - Final	Akkio LAL	20,014,169	BetMGM
11	BetMGM Deterministic Lookalike Audience - Final	Deterministic SQL Scoring	100,000	BetMGM
12	BetMGM Deterministic Lookalike Audience 20M - Final	Deterministic SQL Scoring	20,000,000	BetMGM
13	Holiday Discount Dept Store Shoppers Seed Lookalike - Final	Akkio LAL	7,022,972	Ross, TJMaxx, Marshalls, Burlington, Nordstrom Rack, Target, Walmart, Kohl's, Macy's, JCPenney
14	Holiday Discount Dept Store Shoppers Seed Lookalike 30M - Final	Akkio LAL	30,447,975	Ross, TJMaxx, Marshalls, Burlington, Nordstrom Rack, Target, Walmart, Kohl's, Macy's, JCPenney

Holdout Validation Results

The following results were generated by validating each audience against the **September 2025** transaction holdout window (Sep 1 – Sep 30). Metrics are computed using the same standardized methodology described earlier in this document.

ActBlue Audiences

Metric	LAL (874K)	LAL 20M	Deterministic (500K)	Deterministic 20M
Total Audience IDs	874,141	20,161,266	500,000	20,000,000
Active Matched IDs	228,581	1,496,058	497,399	18,969,993
Brand Shoppers	63,771	64,656	6,948	68,604
Brand Transactions	242,142	243,027	22,051	219,354
Brand Spend	\$5.5M	\$5.6M	\$454K	\$5.0M
Shop Rate	27.90%	4.32%	1.40%	0.36%
Spend Rate	\$24.18	\$3.74	\$0.91	\$0.27
Avg Ticket	\$22.83	\$23.00	\$20.57	\$22.98
Avg Trans / Shopper	3.80	3.76	3.17	3.20
Shop Rate Lift	98.8x	15.3x	4.9x	1.3x
Spend Rate Lift	101.6x	15.7x	3.8x	1.1x

Ross Audiences

Metric	LAL (3.1M)	LAL 30M	Deterministic (500K)	Deterministic 30M
Total Audience IDs	3,075,285	30,831,285	500,000	30,000,000
Active Matched IDs	2,919,651	4,345,506	499,144	28,666,748
Brand Shoppers	1,029,549	1,050,007	82,862	2,286,553
Brand Transactions	1,716,801	1,737,259	130,840	3,529,286

Metric	LAL (3.1M)	LAL 30M	Deterministic (500K)	Deterministic 30M
Brand Spend	\$117.5M	\$118.7M	\$7.5M	\$222.4M
Shop Rate	35.26%	24.16%	16.60%	7.98%
Spend Rate	\$40.25	\$27.32	\$15.08	\$7.76
Avg Ticket	\$68.44	\$68.35	\$57.55	\$63.03
Avg Trans / Shopper	1.67	1.65	1.58	1.54
Shop Rate Lift	5.4x	3.7x	2.6x	1.2x
Spend Rate Lift	6.1x	4.1x	2.3x	1.2x

BetMGM Audiences

Metric	LAL (341K)	LAL 20M	Deterministic (100K)	Deterministic 20M
Total Audience IDs	340,999	20,014,169	100,000	20,000,000
Active Matched IDs	79,769	1,115,457	98,628	18,128,656
Brand Shoppers	25,336	25,728	1,485	29,400
Brand Transactions	371,524	371,916	13,275	227,954
Brand Spend	\$37.0M	\$37.1M	\$874K	\$18.4M
Shop Rate	31.76%	2.31%	1.51%	0.16%
Spend Rate	\$464.34	\$33.24	\$8.86	\$1.02
Avg Ticket	\$99.70	\$99.71	\$65.86	\$80.78
Avg Trans / Shopper	14.66	14.46	8.94	7.75
Shop Rate Lift	239.1x	17.4x	11.3x	1.2x

Metric	LAL (341K)	LAL 20M	Deterministic (100K)	Deterministic 20M
Spend Rate Lift	291.5x	20.9x	5.6x	0.6x

Holiday Discount Department Store Shoppers

Metric	LAL (7.0M)	LAL 30M
Total Audience IDs	7,022,972	30,447,975
Active Matched IDs	6,576,908	7,512,364
Brand Shoppers	5,343,089	5,654,378
Brand Transactions	28,063,998	28,859,924
Brand Spend	\$1.63B	\$1.68B
Shop Rate	81.24%	75.27%
Spend Rate	\$248.25	\$223.78
Avg Ticket	\$58.18	\$58.25
Avg Trans / Shopper	5.25	5.10
Shop Rate Lift	1.7x	1.6x
Spend Rate Lift	2.3x	2.0x

Baseline Reference (All Brands)

Brand	Baseline Active IDs	Baseline Brand Shoppers	Baseline Shop Rate
ActBlue	43,114,357	121,775	0.28%
Ross	43,114,357	2,799,977	6.49%
BetMGM	43,114,357	57,283	0.13%
Holiday Discount Stores	43,114,357	20,446,678	47.42%

Key Observations

Compact Audiences (Precision-Optimized)

ActBlue Akkio LAL (98.8x lift) — The audience concentrates 63,771 political donors out of an 874K audience, achieving a 27.9% shop rate versus a 0.28% general population base rate. The high lift reflects the niche, highly distinctive nature of political donation behavior — donors differ sharply from the general population on both transaction patterns and demographic/interest attributes, making them highly identifiable via lookalike modeling. Average transactions per shopper (3.8) confirms the recurring donation pattern that makes this segment particularly valuable for media targeting.

ActBlue Deterministic LAL (4.9x lift) — The deterministic scoring approach applied to ActBlue delivers a 1.40% shop rate in a 500K audience — 4.9x above the general population baseline. The gap between the Akkio LAL (98.8x) and the deterministic approach (4.9x) reflects the challenge of identifying niche cause/donation behavior through general-purpose RFM + demographic similarity. Political donors don't look distinct in their overall shopping patterns — what distinguishes them is the specific act of donating. That said, the 6,948 donors the deterministic approach identifies are genuine: they average 3.17 transactions per shopper and a \$20.57 average ticket, consistent with ActBlue's recurring small-dollar donation pattern.

Ross Seed LAL (5.4x lift) — With over 1 million brand shoppers in the holdout month out of a 3.1M audience, the 35.3% shop rate at 5.4x lift demonstrates that Ross shopping behavior is highly habitual and the lookalike model effectively identifies repeat shoppers at scale. The spend rate (\$40.25 per active member) represents meaningful media-targetable value.

Ross Deterministic LAL (2.6x lift) — The deterministic SQL-based scoring approach delivers a 16.6% shop rate in a tightly-sized 500K audience. While the lift (2.6x) is more conservative than the Akkio LAL, this audience was built using a fully transparent, reproducible scoring methodology (Gaussian similarity across all RFM + demographic features) with no black-box model.

BetMGM Akkio LAL (239.1x lift) — The standout audience by lift, driven by sports betting's extremely low general population base rate (0.13%). The 31.8% shop rate means nearly 1 in 3 active audience members placed a bet in the holdout month. Average transactions per shopper (14.66) and average ticket (\$99.70) reflect the high-frequency, high-value nature of sports betting — this audience has substantial economic density for activation.

BetMGM Deterministic LAL (11.3x lift) — The deterministic scoring approach applied to BetMGM delivers a 1.51% shop rate in a focused 100K audience — still 11.3x above the general population baseline. While the shop rate is significantly lower than the Akkio LAL (1.51% vs. 31.8%), this reflects the fundamental challenge of identifying a niche behavior (sports betting) through general RFM +

demographic similarity alone. Notably, the bettors identified by the deterministic approach still show high engagement: 8.94 transactions per shopper and a \$65.86 average ticket, confirming these are real, active bettors — just fewer of them.

Holiday Discount Department Store Shoppers (1.7x shop lift, 2.3x spend lift) — The broadest compact audience at 7M IDs, covering a multi-brand retail category (Ross, TJMaxx, Marshalls, Burlington, Nordstrom Rack, Target, Walmart, Kohl's, Macy's, JCPenney). The 81.2% shop rate reflects the ubiquity of discount retail shopping — nearly half the general population already shops these stores in any given month (47.4% baseline). The 2.3x spend lift indicates the audience captures higher-spending shoppers within this high-base-rate category.

Expanded Audiences (Reach-Optimized, 20M–30M)

The expanded audiences reveal the precision/reach degradation curve for each modeling approach — how quickly lift decays as audience size scales from compact to mass-reach. This is the most important comparison for activation planning.

Akkio LAL maintains meaningful lift at scale. The Akkio LAL expanded audiences preserve substantial lift even at 20M–30M:

- **ActBlue LAL 20M (15.3x lift)** — Shop rate drops from 27.9% → 4.3%, but 15.3x lift at 20M is still a highly actionable audience. Notably, the expanded audience captures nearly the same total brand shoppers (64,656 vs. 63,771) — the core ActBlue donor base is almost entirely contained within the compact audience, and the expansion adds reach without diluting the core signal.
- **Ross LAL 30M (3.7x lift)** — Shop rate drops from 35.3% → 24.2%, still well above the 6.5% baseline. The 30M audience captures 1.05M brand shoppers vs. 1.03M in the compact 3.1M — diminishing marginal returns, but the audience remains solidly performant at 10x the size.
- **BetMGM LAL 20M (17.4x lift)** — The most impressive expansion result. Shop rate drops from 31.8% → 2.3%, but 17.4x lift at 20M demonstrates that Akkio's model captured a broad signal around sports betting propensity. Total brand shoppers remain nearly constant (25,728 vs. 25,336), confirming the compact audience already identifies virtually all findable bettors.
- **Holiday LAL 30M (1.6x shop lift, 2.0x spend lift)** — Minimal degradation from the compact audience (81.2% → 75.3% shop rate). The high baseline (47.4%) means even a 30M audience maintains strong absolute shop rates.

Deterministic scoring degrades significantly at scale, but results are brand-dependent. The expanded deterministic audiences show a fundamentally different degradation pattern — lift drops toward 1.0x at 20M–30M, though mainstream retail brands retain modest signal:

- **ActBlue Deterministic 20M (1.3x lift)** — Shop rate drops from 1.40% → 0.36%, barely above the 0.28% baseline. At 20M, general-purpose RFM + demographic similarity cannot meaningfully separate political donors from the population.
- **Ross Deterministic 30M (1.2x lift)** — Shop rate drops from 16.6% → 7.98%, modestly above the 6.49% baseline. The updated population base (`V_AKKIO_ATTRIBUTES_LATEST`) significantly improved panel coverage — active matched IDs increased to 28.7M (95.6% match rate), nearly doubling the brand shoppers found (2.29M) and total brand spend (*222.4M*). *The average ticket* (63.03) remains consistent with real Ross shoppers, and the 1.2x lift demonstrates that deterministic scoring retains a measurable signal for mainstream retail even at 30M scale.
- **BetMGM Deterministic 20M (1.2x shop lift, 0.6x spend lift)** — Shop rate drops to 0.16% vs. 0.13% baseline. The spend rate lift drops below 1.0x, meaning the expanded deterministic audience actually underperforms random selection on spend. This is the clearest evidence that deterministic scoring has a structural ceiling for niche behaviors.

Key takeaway for activation: For campaigns requiring both precision *and* scale, the Akkio LAL expanded audiences are the clear choice — they maintain 15–17x lift at 20M for niche brands and 3.7x at 30M for mainstream retail. The compact deterministic audiences remain valuable for precision-focused campaigns where transparency and reproducibility are priorities. At expanded sizes, deterministic scoring shows a brand-dependent pattern: mainstream retail (Ross at 1.2x) retains modest but measurable lift, while niche behaviors (ActBlue at 1.3x, BetMGM at 1.2x shop / 0.6x spend) hover near or below baseline. The population base expansion to `V_AKKIO_ATTRIBUTES_LATEST` improved the Ross Deterministic 30M results notably — nearly doubling brand shoppers and lifting the shop rate from baseline to 1.2x — but this effect is strongest for high-base-rate categories where demographic and categorical similarity signals remain relevant at scale.

Population Base Impact — Ross Deterministic 30M Re-validation

The Ross Deterministic 30M audience was re-validated after rebuilding on the expanded `V_AKKIO_ATTRIBUTES_LATEST` population base (~50M+ rows) instead of the original `RFM_FEATURES` base (~18M rows). This produced the most significant result change across all fourteen audiences:

Metric	Previous Run	Updated Run	Change
Total Audience IDs	29,969,054	30,000,000	Now reaches full 30M target

Metric	Previous Run	Updated Run	Change
Active Matched IDs	17,302,136	28,666,748	+65.7% — dramatically higher panel coverage
Brand Shoppers	1,163,367	2,286,553	+96.5% — nearly doubled
Brand Transactions	1,793,371	3,529,286	+96.8%
Brand Spend	\$118.7M	\$222.4M	+87.4%
Shop Rate	6.72%	7.98%	+1.26pp — now clearly above the 6.49% baseline
Shop Rate Lift	1.0x	1.2x	From baseline to measurable lift
Spend Rate Lift	1.0x	1.2x	From baseline to measurable lift

Why the improvement matters:

1. **Panel coverage was the bottleneck, not scoring quality.** The previous run's 57.7% active match rate (17.3M/30M) meant nearly half the audience had no observable transactions — inflating the denominator with unmeasurable members and depressing the apparent shop rate. The updated run's 95.6% match rate (28.7M/30M) provides a far more complete measurement surface.
2. **The scoring signal was always present — it just needed a larger population to express itself.** The expanded population base ensures the top 30M scores are drawn from a 50M+ candidate pool rather than exhausting an 18M pool. Members ranked 18M–30M under the old population had no RFM data and scored on categorical features alone; under the new population, those ranks are filled by members with richer feature profiles.
3. **Brand-dependent effect.** Ross's high base rate (6.49%) means even modest scoring improvements translate to observable lift. For niche brands (ActBlue at 0.28%, BetMGM at 0.13%), the same population expansion would not produce comparable improvement because the target behavior is too rare for demographic/categorical similarity to capture at scale.
4. **Spend-side consistency confirms audience quality.** The average ticket held steady (66.19 → 63.03), and transactions per shopper remained constant (1.54), indicating the newly captured shoppers exhibit the same purchase patterns as the original cohort — not a lower-quality tail.

Comparison to AFS Purchase Propensity Modeling Approach

AFS employs a structured ML propensity modeling pipeline with a strict temporal holdout design:

Step	AFS Propensity Approach	Akkio Lookalike Approach
Label creation	Binary classification — shoppers (positive) vs. non-shoppers (negative) identified from Sept 2024 – Sept 2025	Seed identification — brand shoppers identified from transaction data with quality filters
Feature engineering	Historical features from Sept 2023 – Sept 2024 (1-year temporal gap between features and labels)	Pre-materialized RFM features (5 time windows) + 50+ demographic/interest/propensity attributes
Modeling	Two separate models: demographics-only and RFM-only; evaluated via AUC, Lift, and top-decile precision	Gaussian similarity scoring weighted by seed-vs-population divergence (deterministic), or Akkio's ML platform (LAL)
Output	Probability-ranked audience segments (top 1%, 5%, 10%)	Similarity-ranked audience with configurable size cutoff
Temporal separation	1-year gap between feature window and label window	Features and seeds use overlapping data (see note below)

The two approaches are **complementary**:

- **AFS's propensity model** uses a rigorous 1-year temporal separation between features and labels, which produces unbiased estimates of predictive power. The separate demographics-only and RFM-only models allow attribution of which feature classes drive the most signal per brand.
- **The Akkio/deterministic approach** incorporates a broader feature set in a single unified score (RFM across 5 windows + all demographics + interests + propensities), which can capture complex multi-factor patterns. The trade-off is that the current validation cycle has some temporal overlap between features/seeds and the holdout window (see below).

Methodology update for expanded audiences: The deterministic scoring pipeline was updated to use `V_AKKIO_ATTRIBUTES_LATEST` (~50M+ rows) as the population base instead of `RFM_FEATURES` (~18M rows). This change was necessary to support audience sizes of 20M–30M. Members without RFM data score 0 on numeric features (via existing COALESCE logic) but still contribute categorical

similarity scores. The ranking naturally prioritizes members with transaction history, so the top of the expanded audiences is behaviorally rich while the tail extends into the demographics-only population. This population base change is reflected in the updated `affinity_client_context.md` and the deterministic lookalike SQL templates.

For the Oct-Dec validation window, both approaches will be evaluated under identical conditions — a true out-of-sample test where neither approach had access to the holdout data. This will provide a clean, apples-to-apples comparison of audience quality across both compact and expanded audience sizes.

Validation Caveats & Oct-Dec Expectations

Temporal overlap in current validation: The September holdout validation presented above has a known limitation — the RFM features and seed identification both use data through end of September 2025, which overlaps with the September holdout window. This means:

1. Seed members identified from September transactions are captured in the lookalike → they will naturally appear as brand shoppers in the September holdout (circular)
2. RFM features (especially 1-month and 3-month windows) incorporate September behavior, giving the scoring model information about the holdout period

As a result, the shop rates and lifts above are **optimistic** relative to a clean out-of-sample test. They demonstrate that the audiences are correctly constructed and enriched for the target behaviors, but the exact magnitudes should be interpreted as upper bounds.

Why we remain confident for Oct-Dec:

1. **Behavioral persistence** — The brands in this portfolio exhibit highly habitual consumer behavior. Political donors (ActBlue) give repeatedly. Retail shoppers (Ross) at 3+ transactions per year are habitual visitors. Sports bettors (BetMGM) at 14.7 transactions/month are deeply engaged. These patterns persist month-over-month and are not artifacts of a single month's data.
2. **Lift headroom** — Even with a significant discount for temporal overlap, the lift numbers have substantial margin. ActBlue at 99x and BetMGM at 239x could lose 50-70% of their lift and still represent exceptional audiences. Ross at 5.4x could halve and remain a strong 2-3x performer.
3. **Behavioral signatures pass sanity checks** — Average ticket values are consistent with brand economics (Ross ~\$68, *ActBlue* 23, BetMGM ~\$100). Transactions per shopper reflect real behavioral patterns (3.8 recurring donations vs. 1.6 retail visits vs. 14.7 betting sessions). These would not be coherent if the audiences were purely artifacts of leakage.

4. **AFS's Oct-Dec validation is the definitive test** — The October through December window was not available during any stage of audience construction (seed identification, feature engineering, or scoring). This will provide fully unbiased metrics against which to measure audience quality and compare approaches.

Expected Oct-Dec performance ranges:

Audience	Sept Shop Rate	Expected Oct-Dec Range	Rationale
ActBlue LAL (874K)	27.9%	12-20%	Recurring donors persist; some Sept-only donors drop off
ActBlue LAL 20M	4.3%	2-3%	Expanded audience maintains structure; donor persistence provides floor
ActBlue Deterministic (500K)	1.40%	0.5-1.0%	Niche cause behavior harder to capture deterministically; recurring donors provide baseline lift
ActBlue Deterministic 20M	0.36%	~0.28%	Near baseline; limited incremental value expected
Ross LAL (3.1M)	35.3%	15-25%	Habitual retail; holiday season may boost or maintain
Ross LAL 30M	24.2%	10-18%	Broad audience; holiday season tailwind helps
Ross Deterministic (500K)	16.6%	8-14%	More conservative baseline; most honest current signal
Ross Deterministic 30M	8.0%	6.5-7.5%	Modest lift (1.2x) now observed; habitual retail behavior may sustain signal
BetMGM LAL (341K)	31.8%	15-25%	NFL season Oct-Dec is peak betting; strong tailwind
BetMGM LAL 20M	2.3%	1-2%	Peak NFL season may partially offset expansion dilution
BetMGM Deterministic	1.51%	0.5-1.2%	Niche behavior harder to capture deterministically; still meaningful lift

Audience	Sept Shop Rate	Expected Oct-Dec Range	Rationale
(100K)			expected
BetMGM Deterministic 20M	0.16%	~0.13%	At baseline; no meaningful lift expected
Holiday Discount Store LAL (7M)	81.2%	65-80%	Holiday season is the target period; baseline also rises
Holiday Discount Store LAL 30M	75.3%	60-75%	Minimal degradation expected; category ubiquity at scale

Delivery Status

All fourteen final audiences have been delivered to the AFS S3 bucket for activation and validation. Each audience file contains the audience member AFS_IDs as scored and ranked by the respective modeling approach. For each brand, both a compact (precision-optimized) and expanded (reach-optimized, 20M–30M) variant are available, enabling activation teams to select the appropriate audience based on campaign scale and performance requirements.

How to Interpret These Results

Shop Rate Is the Primary Quality Signal

Shop Rate answers the core question: *"What percentage of active audience members actually purchased from the brand during the holdout window?"*

- **Higher shop rate = better audience quality** — the audience is more concentrated with likely buyers
- **Seed audiences will always have higher shop rates** — they are built from known buyers, so holdout shop rate measures retention, not prediction
- **Lookalike / propensity audiences are the true test** — their shop rate reflects the model's ability to identify *new* buyers

Spend Rate Captures Dollar-Value Impact

Spend Rate combines shop rate with purchase value to answer: *"How much brand revenue does each active audience member represent?"*

This is the most actionable metric for media planning — it directly translates audience quality to economic value.

Lift Over Baseline

To contextualize these results, compare against a **random baseline** — the shop rate you would observe by taking a random sample of the same size from the active population. Any audience's shop rate divided by the baseline shop rate gives you **lift**:

$$\text{Lift} = \text{Audience Shop Rate} / \text{Random Baseline Shop Rate}$$

A lift of 2.0x means the audience is twice as likely to contain brand shoppers as a random sample. Higher lift = better targeting precision.

Summary

The audience validation framework provides a rigorous, transparent method for measuring audience quality:

1. **Temporal holdout design** ensures metrics reflect true predictive power by evaluating audiences against unseen future transaction data
2. **Standardized metrics** (Shop Rate, Spend Rate, Avg Ticket, Lift) enable direct comparison across audiences, brands, and modeling approaches
3. **Active Matched IDs denominator** ensures fair comparison regardless of panel coverage differences
4. **Baseline lift comparison** quantifies how much better each audience performs relative to a random sample of the general population
5. **Cross-methodology comparison** enables evaluation of Akkio's ML-based lookalike modeling alongside deterministic similarity scoring and AFS's propensity modeling approach
6. **Precision/reach curve analysis** — the expanded (20M–30M) audiences reveal how each approach degrades at scale, providing activation teams with clear guidance on audience size selection

All fourteen delivered audiences (seven compact, seven expanded) span four brands across two construction methods. The compact audiences demonstrate meaningful lift for all brands and methods, with the strongest results for niche/cause brands (ActBlue at 98.8x, BetMGM at 239.1x). The expanded audiences reveal a critical finding: **Akkio LAL maintains 15–17x lift for niche brands and 3.7x for mainstream retail at 20M–30M scale**, while deterministic scoring shows brand-dependent degradation — dropping to near-baseline for niche behaviors (ActBlue 1.3x, BetMGM 1.2x) but retaining modest lift for mainstream retail (Ross 1.2x) following the population base expansion to `V_AKKIO_ATTRIBUTES_LATEST`. This makes the Akkio LAL the preferred approach for campaigns requiring both quality and scale, though expanded deterministic audiences remain viable for high-base-rate retail categories. AFS's upcoming October – December validation will provide the definitive out-of-sample test of audience quality across all fourteen audiences.

Appendix: Lookalike Audience Design Rationale

The following sections provide detailed design rationale and reference SQL for the deterministic lookalike methodology. The compact rules for query generation live in `affinity_client_context.md`; this appendix explains the *why* behind those rules.

Why Scoring Beats Thresholds

Binary pass/fail thresholds on a limited feature set (e.g., RFM-only) have three structural problems:

1. **No gradient** — A prospect who barely misses one threshold is treated the same as one who misses everything. There is no "how similar" signal.
2. **Limited signal** — Using only RFM for matching discards the strongest predictors for many brands. For niche/cause brands (e.g., ActBlue), demographics and interests are often more discriminative than general transaction behavior. For retail brands, channel and category affinity may matter as much as spend.
3. **No ranking** — Everyone who clears the thresholds is "equally good," with no way to prioritize the best prospects or control audience size precisely.

A scoring approach solves all three: every feature contributes a weighted signal, prospects are ranked by total similarity, and audience size is controlled by taking the top N. Seed members naturally score high (they match the seed profile by definition) without being force-included, producing honest validation metrics.

Seed Member Handling

Seed members are included in the scored population and are NOT force-excluded. They naturally rank near the top because they match the seed profile. This mirrors how propensity models work:

- A propensity model scores everyone; known buyers score well organically
- Force-excluding seed creates an artificial gap and can bias the LAL toward lower-quality prospects
- Force-including seed (as a union) inflates validation metrics without reflecting LAL quality

When validating, segment results by seed vs. non-seed to measure incremental LAL lift separately. The delivered audience should contain seed members naturally ranked by score.

Precision vs. Reach Tuning

Audience size (<audience_size>) directly controls the precision/reach trade-off:

Goal	Audience Size	Expected Outcome
High Precision	Small (top 50K–500K)	Highest-scoring prospects; strong similarity, highest expected conversion
Balanced	Medium (top 500K–3M)	Good similarity with broader reach; moderate conversion
High Reach	Large (top 20M–30M)	Wider net; scores taper, conversion approaches baseline at the margin

The score distribution itself is informative: a steep drop-off means the seed is highly distinctive and a tight audience is appropriate. A gradual decline means the seed blends with the population and larger audiences are needed.

The expanded audience validation results (see Holdout Validation Results) confirm this gradient empirically: Akkio LAL audiences maintain meaningful lift (3.7–17.4x) at 20M–30M, while deterministic scoring drops to 1.2–1.3x at the same sizes. The degradation is brand-dependent — mainstream retail (Ross at 1.2x) retains modest signal, while niche behaviors converge toward baseline. Akkio LAL remains the preferred method when both quality and scale are required.

Small Seed Safeguards — Detailed Explanation

Niche brands (political donations, sports betting, luxury goods) often produce very small seeds — sometimes fewer than 10 people. The scoring methodology handles this through two critical floors:

Bandwidth Floor: `GREATEST(COALESCE(seed_std, 0), 0.5 * pop_std)`

- Seed size = 1: `STDDEV()` returns `NULL` → falls back to `0.5 * pop_std`
- Seed size = 2–10 with similar values: `STDDEV()` near-zero → Gaussian becomes needle-thin → falls back to `0.5 * pop_std`
- Normal seed variance: `seed_std > 0.5 * pop_std` → uses `seed_std` (no change)
- The 0.5 multiplier means the Gaussian is twice as selective as the population spread — still providing meaningful differentiation while preventing collapse

Importance Floor: `GREATEST(importance, 0.1)`

- Very large/common seeds (e.g., "all holiday shoppers"): `Seed ≈ population` → `importance ≈ 0` → floor ensures Gaussian still differentiates by proximity to seed mean

Seed Count Diagnostic: Always output seed count. If seed is empty (0 members), return informative message rather than misleading all-zero audience.

Key Design Points

- **V_AKKIO_ATTRIBUTES_LATEST as population base:** The scoring pipeline uses `V_AKKIO_ATTRIBUTES_LATEST` (~50M+ rows) as the base table, `LEFT JOIN`d to `RFM_FEATURES` for behavioral features. This ensures audience sizes up to 30M+ can be satisfied. Members without RFM data score 0 on numeric features but still contribute categorical similarity — the ranking naturally prioritizes behaviorally-rich members.
- **Pre-materialized RFM via RFM_FEATURES :** Scoring reads pre-computed features — no transaction-level aggregation at scoring time, eliminating the biggest performance bottleneck.
- **Single CTE chain:** The entire query is one `CREATE TABLE AS WITH ... SELECT ...` statement. No temp tables. Since `RFM_FEATURES` eliminates the heavy scan, a CTE chain executes efficiently.
- **Pre-computed categorical importance:** Each `IMP_*` CTE computes the scalar importance once. The scoring `SELECT` references these scalars instead of re-evaluating correlated subqueries per row.
- **No expression duplication:** `NUMERIC_SIMILARITY_SCORE` and `CATEGORICAL_SIMILARITY_SCORE` are computed once in the `SCORED` CTE, then summed in the final `SELECT`.
- **Gaussian similarity** ($\text{EXP}(-0.5 * z^2)$) produces a smooth 0-to-1 score per feature. A prospect matching the seed mean exactly scores 1.0; distant values taper toward 0.
- **Automatic feature weighting** via `importance = |seed_mean - pop_mean| / pop_stddev` — no manual weight tuning needed.

- **All features contribute** — RFM (all 5 windows), demographics, interests, propensities. The weighting ensures discriminative features matter most.

Reference SQL — Full Lookalike Scoring Query

The following is the complete SQL pattern for a lookalike audience build. The compact version in `affinity_client_context.md` describes the CTE structure; this is the fully expanded reference.

```

CREATE TABLE <output_table> AS
WITH
-- Phase 1: Seed identification & population features
SEED_IDS AS (
    SELECT DISTINCT AKKIO_ID
    FROM FACT_TRANSACTION_ENRICHED
    WHERE (<brand_filter>)
        AND TRANS_DATE >= DATEADD(MONTH, -12, '<ref_date>'::DATE)
        AND TRANS_DATE < '<cutoff_day_after>'
),

POP_FEATURES AS (
    SELECT
        d.AKKIO_ID,
        r.days_since_last_txn,
        r.tot_trans_12mo, r.tot_spend_12mo,
        r.tot_online_trans_12mo, r.tot_online_spend_12mo,
        r.avg_days_btwn_trans_12mo, r.brand_diversity_12mo, r.online_ratio_12mo,
        r.tot_trans_9mo, r.tot_spend_9mo, r.tot_online_trans_9mo, r.tot_online_spend_9mo,
        r.avg_days_btwn_trans_9mo, r.brand_diversity_9mo,
        r.tot_trans_6mo, r.tot_spend_6mo, r.tot_online_trans_6mo, r.tot_online_spend_6mo,
        r.avg_days_btwn_trans_6mo, r.brand_diversity_6mo,
        r.tot_trans_3mo, r.tot_spend_3mo, r.tot_online_trans_3mo, r.tot_online_spend_3mo,
        r.avg_days_btwn_trans_3mo, r.brand_diversity_3mo,
        r.tot_trans_1mo, r.tot_spend_1mo, r.tot_online_trans_1mo, r.tot_online_spend_1mo,
        r.avg_days_btwn_trans_1mo, r.brand_diversity_1mo,
        CASE WHEN s.AKKIO_ID IS NOT NULL THEN 1 ELSE 0 END AS IS_SEED,
        d.GENDER, d.STATE, d.POLITICS, d.INCOME_BUCKET, d.EDUCATION_LEVEL,
        d.ETHNICITY, d.AGE, d.MARITAL_STATUS, d.HOMEOWNER_STATUS,
        d.NET_WORTH_BUCKET, d.OCCUPATION
    -- ... include ALL demographic, interest, and propensity fields from V_AKKIO_ATTRIB
    FROM V_AKKIO_ATTRIBUTES_LATEST d
    LEFT JOIN RFM_FEATURES r ON d.AKKIO_ID = r.AKKIO_ID
    LEFT JOIN SEED_IDS s ON d.AKKIO_ID = s.AKKIO_ID
),

-- Phase 2: Statistics & scoring
SEED_NUMERIC_STATS AS (
    SELECT
        AVG(days_since_last_txn) AS seed_mean_recency,
        STDDEV(days_since_last_txn) AS seed_std_recency,
        AVG(tot_trans_12mo) AS seed_mean_freq,
        STDDEV(tot_trans_12mo) AS seed_std_freq,

```

```

    AVG(tot_spend_12mo) AS seed_mean_spend,
    STDDEV(tot_spend_12mo) AS seed_std_spend,
    AVG(tot_trans_3mo) AS seed_mean_freq_3mo,
    STDDEV(tot_trans_3mo) AS seed_std_freq_3mo,
    AVG(online_ratio_12mo) AS seed_mean_online_ratio,
    STDDEV(online_ratio_12mo) AS seed_std_online_ratio,
    AVG(brand_diversity_12mo) AS seed_mean_brand_div,
    STDDEV(brand_diversity_12mo) AS seed_std_brand_div,
    AVG(avg_days_btwn_trans_12mo) AS seed_mean_cadence,
    STDDEV(avg_days_btwn_trans_12mo) AS seed_std_cadence,
    AVG(CAST(AGE AS FLOAT)) AS seed_mean_age,
    STDDEV(CAST(AGE AS FLOAT)) AS seed_std_age
    -- ... repeat for ALL numeric features
FROM POP_FEATURES
WHERE IS_SEED = 1
),

```

```

POP_NUMERIC_STATS AS (
    SELECT
        AVG(days_since_last_txn) AS pop_mean_recency,
        STDDEV(days_since_last_txn) AS pop_std_recency,
        AVG(tot_trans_12mo) AS pop_mean_freq,
        STDDEV(tot_trans_12mo) AS pop_std_freq,
        AVG(tot_spend_12mo) AS pop_mean_spend,
        STDDEV(tot_spend_12mo) AS pop_std_spend,
        AVG(tot_trans_3mo) AS pop_mean_freq_3mo,
        STDDEV(tot_trans_3mo) AS pop_std_freq_3mo,
        AVG(online_ratio_12mo) AS pop_mean_online_ratio,
        STDDEV(online_ratio_12mo) AS pop_std_online_ratio,
        AVG(brand_diversity_12mo) AS pop_mean_brand_div,
        STDDEV(brand_diversity_12mo) AS pop_std_brand_div,
        AVG(avg_days_btwn_trans_12mo) AS pop_mean_cadence,
        STDDEV(avg_days_btwn_trans_12mo) AS pop_std_cadence,
        AVG(CAST(AGE AS FLOAT)) AS pop_mean_age,
        STDDEV(CAST(AGE AS FLOAT)) AS pop_std_age
        -- ... same fields as SEED_NUMERIC_STATS
    FROM POP_FEATURES
),

```

```

-- Categorical distribution + importance (repeat pattern for every categorical field)
SEED_CAT_GENDER AS (
    SELECT GENDER AS cat_value, COUNT(*)::FLOAT / SUM(COUNT(*)) OVER () AS seed_share
    FROM POP_FEATURES WHERE IS_SEED = 1 AND GENDER IS NOT NULL GROUP BY GENDER

```



```

),
POP_CAT_GENDER AS (
    SELECT GENDER AS cat_value, COUNT(*)::FLOAT / SUM(COUNT(*)) OVER () AS pop_share
    FROM POP_FEATURES WHERE GENDER IS NOT NULL GROUP BY GENDER
),
IMP_GENDER AS (
    SELECT GREATEST(MAX(sc.seed_share / NULLIF(pc.pop_share, 0)), 0.1) AS importance
    FROM SEED_CAT_GENDER sc JOIN POP_CAT_GENDER pc ON sc.cat_value = pc.cat_value
),
-- ... repeat SEED_CAT_, POP_CAT_, IMP_ for: STATE, POLITICS, INCOME_BUCKET,
--     EDUCATION_LEVEL, ETHNICITY, MARITAL_STATUS, HOMEOWNER_STATUS,
--     NET_WORTH_BUCKET, OCCUPATION, and ALL other categorical fields

SCORED AS (
    SELECT
        P.AKKIO_ID,
        P.IS_SEED,

        -- Numeric:  $\exp(-0.5 * ((val - seed\_mean) / bandwidth)^2) * importance$ 
        EXP(-0.5 * POW((P.days_since_last_txn - S.seed_mean_recency)
            / GREATEST(COALESCE(S.seed_std_recency, 0), 0.5 * POP.pop_std_recency), 2))
        * GREATEST(ABS(S.seed_mean_recency - POP.pop_mean_recency) / NULLIF(POP.pop_std_r
+ EXP(-0.5 * POW((P.tot_trans_12mo - S.seed_mean_freq)
            / GREATEST(COALESCE(S.seed_std_freq, 0), 0.5 * POP.pop_std_freq), 2))
        * GREATEST(ABS(S.seed_mean_freq - POP.pop_mean_freq) / NULLIF(POP.pop_std_freq, 0
+ EXP(-0.5 * POW((P.tot_spend_12mo - S.seed_mean_spend)
            / GREATEST(COALESCE(S.seed_std_spend, 0), 0.5 * POP.pop_std_spend), 2))
        * GREATEST(ABS(S.seed_mean_spend - POP.pop_mean_spend) / NULLIF(POP.pop_std_spend
-- + ... repeat for ALL other numeric features
        AS NUMERIC_SIMILARITY_SCORE,

        -- Categorical: seed_share * scalar importance from IMP_ CTEs
        COALESCE(SG.seed_share, 0) * (SELECT importance FROM IMP_GENDER)
        + COALESCE(SS.seed_share, 0) * (SELECT importance FROM IMP_STATE)
        -- + ... repeat for ALL other categorical fields
        AS CATEGORICAL_SIMILARITY_SCORE

    FROM POP_FEATURES P
    CROSS JOIN SEED_NUMERIC_STATS S
    CROSS JOIN POP_NUMERIC_STATS POP
    LEFT JOIN SEED_CAT_GENDER SG ON P.GENDER = SG.cat_value
    LEFT JOIN SEED_CAT_STATE SS ON P.STATE = SS.cat_value
    -- ... LEFT JOIN for each categorical seed distribution CTE

```

```
)

-- Phase 3: Rank and extract
SELECT
    AKKIO_ID, IS_SEED,
    NUMERIC_SIMILARITY_SCORE, CATEGORICAL_SIMILARITY_SCORE,
    (NUMERIC_SIMILARITY_SCORE + CATEGORICAL_SIMILARITY_SCORE) AS SIMILARITY_SCORE
FROM SCORED
ORDER BY SIMILARITY_SCORE DESC
LIMIT <audience_size>;
```