



white cup **black lid**
white paper

Signal Engine: White Paper

Published June 2026

Leading Indicators in Residential Real Estate

How public signal data can give broker teams an earlier read on listing supply, market momentum, and buyer demand

Most market intelligence tools available to broker teams are retrospective — they describe what the MLS already recorded. This paper examines whether a class of publicly available signals (search interest, news and employer events, listing portal behavior) can function as leading indicators: patterns that precede measurable market movement rather than confirm it. The objective is not to claim predictive accuracy, but to identify which signals are worth monitoring, what they might indicate in advance, and what validation would be required before treating them as actionable.

Section 1

The information gap most broker teams don't measure

Every major decision a broker team makes — which neighborhoods to farm, when to pitch a listing, how to price a home, where to focus agent effort this month — is made against a picture of the market that is, by definition, already out of date.

The MLS is a recording. It documents transactions, price changes, and inventory levels after they occur. The gap between when a seller begins thinking about moving and when a listing appears in the MLS can be weeks or months. In that window, a seller is reachable, not yet committed to another agent — but invisible to any tool that relies on MLS data as its primary input.

This paper examines a different question: are there publicly observable signals that tend to precede listing activity, price movement, or buyer demand spikes — and if so, what would it take to monitor them systematically?

To be clear about what this is not: this is not a claim that such signals reliably predict market outcomes. The honest answer is that we do not yet know whether they do, for any given market. The purpose of this paper is to lay out what the candidate signals are, what prior evidence (from adjacent fields) suggests about their behavior, and what a disciplined test would look like in a real market.

A note on terminology

Throughout this paper, "leading indicator" means a signal that has historically preceded a measurable outcome by a meaningful interval — not simply a signal that correlates with or confirms an outcome already visible in the MLS. We use the term carefully, because the distinction matters for whether a broker team can act on the information in time to change what they do.

Section 2

Candidate signals and what they may indicate

Four categories of publicly observable data appear in the research literature and in practitioner accounts as potentially leading real estate market activity. Each has a different data availability profile, a different interpretive challenge, and a different relationship to the outcomes brokers most care about.

Search interest — the earliest observable intent signal

When a homeowner begins to consider selling, one of their first actions is typically a search. Queries like "what is my home worth," "sell my house [city]," or "home values in [neighborhood]" precede any contact with an agent, any call to a brokerage, and almost always precede any listing agreement. These queries are observable in aggregate through tools like Google Trends, which reports relative search volume by geography and time period.

The potential value here is timing. A geographic cluster of seller-intent searches may indicate that a cohort of homeowners in that area is entering the consideration phase simultaneously — possibly triggered by a shared local event (a neighborhood sale at an unexpected price, a news story about appreciation, a rate change). If that cohort proceeds to list, the search spike would have preceded the inventory increase by some interval.

The interpretive challenge is that search volume is noisy. Seasonal patterns, media cycles, and general curiosity all drive searches that never convert to listings. Whether seller-intent search spikes in a specific geography are predictive of inventory movement in that geography — and over what time horizon — is an empirical question, not one that can be answered from search data alone. It requires a ground-truth comparison against actual listing data.

The analogous use case in music analytics offers a partial precedent: artist-name search spikes on platforms like Spotify have been shown to precede streaming catalog activity in certain contexts. The mechanism is similar — search captures intent before the transaction occurs — but the lag structure and the signal-to-noise ratio differ enough that the music finding cannot be directly transplanted to real estate without independent validation.

News and employer events — supply-demand shocks with a timestamp

A different class of signals operates through economic fundamentals rather than individual intent. Large employer announcements — a new corporate headquarters locating to a metro, a major employer announcing layoffs, a relocation incentive program — predictably alter migration patterns and housing demand. Unlike search interest, these events are discrete and dated, and their directional effect on local housing demand is generally well understood.

The practical question for a broker team is whether local news monitoring can surface these events early enough to be useful. In many cases, employer announcements precede any visible shift in MLS activity by six to eighteen months. A team that identifies an inbound employer event and begins positioning in the relevant submarkets before buyer demand materializes has a genuine lead-time advantage — one that does not require any proprietary data.

Listing portal behavior — demand-side pressure, partially observable

Listing portals (Zillow, Realtor.com, Redfin) publish certain aggregate metrics that capture buyer-side demand: listing view counts, save rates, and days-on-market statistics. These are not leading indicators in the strict sense — they reflect activity on listings already in the market — but they can indicate whether a submarket is absorbing inventory quickly or slowly, and whether buyer competition is increasing or decreasing in a given area.

When read alongside search interest data, portal behavior can provide a more complete picture: search intent is the leading edge, and portal engagement is confirmation that interest is translating into active buyer activity. Divergence between the two — high search interest but low portal engagement, for example — may indicate a submarket where sellers are actively being researched but available inventory has not yet surfaced to meet the demand.

Structural market indicators — confirming, not leading

Days-on-market compression, price-cut frequency, and listing volume trends are the signals most broker teams already track. They are important, but they occupy a different position in the information sequence: they confirm that a market is moving, rather than indicating that it is about to move.

Their value in a signal-monitoring framework is as validators. If search interest rises in a given ZIP and is followed, some weeks later, by DOM compression and a reduction in price

cuts, that sequence provides evidence that the leading signal was in fact predictive. If the structural indicators do not follow, the leading signal was noise. Over time, this comparison is how predictive validity — or its absence — gets established.

Summary: signal positions in the information sequence

Position	Signal type
Earlier (potentially leading)	Seller-intent search queries; employer and relocation news
Concurrent (demand confirmation)	Portal view and save rates; buyer-intent search
Lagging (outcome confirmation)	Days-on-market, price cuts, listing volume, closings

The practical question is whether the earlier signals are predictive — in a specific market, over a specific time window — or simply correlated. That cannot be determined without a ground-truth comparison against actual MLS data.

Section 3

Which outcomes are worth trying to forecast

Not every market outcome a broker team cares about is equally amenable to signal-based forecasting. Some are structurally harder — they require granular data that is not publicly available, or they operate on time horizons too short to act on. Others are well-matched to the signals described in Section 2. The framework below ranks candidate outcomes by two dimensions: how much a broker team benefits from advance knowledge, and how realistic it is to construct a leading indicator using publicly available data.

Outcome	What a forecast would tell you	Feasibility with public data
Listing inventory surge in a farm area	A ZIP or neighborhood is likely to produce above-average new listings in the next 4–8 weeks	Moderate. Seller-intent search and equity data are available; lag structure must be validated per market.
Market regime shift (heating / cooling)	A submarket is at or approaching a directional inflection — moving from cooling to heating or vice versa	Moderate–High. Composite of search, DOM, and price-cut data; more confirmatory than predictive.
Buyer demand spike by geography	Buyer competition in a specific ZIP is increasing, driven by inbound migration or employer events	High for employer events (discrete, dated). Moderate for diffuse migration signals.
Migration flow by metro or ZIP	Net inflow or outflow from a geography, before it is reflected in MLS transaction volume	Moderate. Search-based migration proxies exist; precision at ZIP level is limited.
Per-listing price-cut or stale risk	A specific active listing is at elevated risk of needing a price reduction or missing its DOM target	Low with public data alone. Requires access to the broker's own listing feed as input.

Table 1. Candidate forecast outcomes ranked by broker-team value and public-data feasibility.

The listing inventory question deserves particular attention

Of the outcomes listed above, the listing inventory surge forecast is simultaneously the most valuable to a broker team and the most methodologically demanding to build well. Listing supply — not buyer demand — is the scarce resource in most competitive residential

markets. The teams that win a disproportionate share of listings are typically the ones that reach sellers earliest, before the seller has contacted multiple brokerages.

The hypothesis is that seller-intent search behavior, combined with local equity appreciation data, can identify geographic clusters of homeowners who are "in the window" — actively considering a move but not yet listed. If that hypothesis holds in a given market, the lead time between signal and listing would be the period during which the broker team can act differently than competitors relying on MLS data alone.

The hypothesis is not proven. It is worth testing.

Section 4

Methodology and what honest validation requires

The history of data-driven forecasting in real estate contains a recurring failure mode: a signal that correlates well with past outcomes is presented as predictive of future ones, without the structural test that distinguishes correlation from genuine leading-indicator status. The AVM (automated valuation model) literature is full of examples. So is the agent-productivity scoring space.

This matters for broker teams evaluating any signal-based platform, including the one described in this paper. Before treating a signal as actionable, three tests should be satisfied:

1. Pre-registration

The forecast — the specific signal, the specific outcome, the specific geography, and the predicted time window — must be committed to before the outcome is observed. A forecast that is adjusted after the fact to fit what happened is not a forecast; it is a retrospective explanation. Any credible pilot should establish a locked prediction before the measurement window opens.

2. Ground-truth comparison

The predicted outcome must be compared against independently recorded data — ideally the broker's own MLS feed — not against further signal data. If a search interest spike is "validated" by reference to other search data, nothing has been proven. Validation requires a comparison to the actual outcome the signal was supposed to predict: did inventory in that ZIP increase in that window?

3. Out-of-sample evaluation

A model that was tuned on historical data from one market and then applied to a different market, or a different time period, without re-validation, is unreliable. Lead-time structures differ across markets, price tiers, and economic conditions. The only honest accuracy number is prospective — a locked geography and window, evaluated after the fact.

What accurate claims look like in practice

A defensible claim is not: "Our platform predicts when neighborhoods will list." A defensible claim is: "In our pilot markets, seller-intent search spikes in a given ZIP were followed by above-average new listing volume within 6 weeks in 7 of 10 observed instances." The second claim specifies a market, an outcome, a window, and a hit rate. The first claim specifies none of those things, and cannot be evaluated.

What we are not building, and why that matters

Several adjacent products are explicitly outside the scope of signal-based market monitoring as described here. Naming them clearly is useful, because they represent categories that broker teams sometimes request and that carry different technical, legal, and ethical profiles.

Individual homeowner targeting

Parcel-level prediction of which specific homeowners are likely to sell is a different product from neighborhood-level signal monitoring. It requires different data (often purchased from data brokers), operates under different privacy considerations, and is subject to fair housing compliance scrutiny. This paper addresses only neighborhood-level and ZIP-level signal patterns from public data.

Automated valuation

AVM products attempt to estimate the current market value of a specific property. That is a distinct modeling problem from forecasting whether a neighborhood is likely to see increased listing activity. The two products share some inputs (price appreciation data) but have entirely different outputs, validation standards, and liability profiles.

Agent productivity scoring

Some platforms attempt to predict which agents are likely to win listings or close transactions based on historical performance data. This is a CRM and operations problem, not a market signal problem, and is outside the scope of external-signal monitoring.

Section 5

What a structured pilot would look like

A pilot designed to test the hypothesis described in Section 3 — that seller-intent search behavior can predict listing inventory surges — would need to satisfy the methodological standards described in Section 4 to be worth running. The following is a description of a pilot structure that meets those standards.

Step 1: Define the geography

The broker team identifies three to five ZIP codes or named farm areas to monitor. These should represent a meaningful portion of the team's current or target territory — not cherry-picked areas where activity is already known to be elevated.

Step 2: Select and lock the forecast

Before the pilot begins, a specific outcome is defined: for example, "above-baseline new listing volume in the target ZIPs within a 6-week window following a search interest threshold." That definition does not change once the pilot starts. The signal thresholds and the outcome criteria are both committed to in writing before any data is collected.

Step 3: Monitor and alert

The platform monitors the defined signals across the target geographies and generates alerts when a signal pattern crosses the defined threshold. The broker team receives these alerts and can observe whether they align with their own market knowledge — but neither confirms nor adjusts the forecast based on that observation during the pilot.

Step 4: Evaluate against MLS ground-truth

At the close of the pilot window (typically 60–90 days), the predicted outcomes are compared to what actually happened in the MLS. The evaluation is binary: did the target ZIP see above-baseline listing volume in the defined window following the signal? The hit rate across the target geographies is the accuracy measure.

Step 5: Report findings honestly

If the signals were predictive, the pilot report describes the hit rate, the observed lead time, and the conditions under which the signals fired. If the signals were not predictive — or if the hit rate was within the margin of chance — the report says that, along with the most likely explanations.

On the question of MLS data sharing

Ground-truth validation is substantially easier if the broker team is willing to share anonymized MLS data — specifically, listing volume by ZIP and date — for the pilot period. This data is used only as a validation input, not as a product input; the signal monitoring itself relies entirely on public data. A pilot can proceed without MLS access, but validation will be slower and require manual comparison against publicly reported market statistics.

Section 6

Open questions this paper does not resolve

A white paper written before a hypothesis has been tested should be honest about what it does not know. The following are the questions that the approach described here cannot answer without pilot data from a real market.

Does the lead time exist — and if so, how long is it?

Whether seller-intent search in a given ZIP precedes listing activity at all, and if it does, by how many weeks, is unknown until tested in a specific market. The value of any leading indicator is directly proportional to the lead time it provides. A signal that arrives two days before a listing hits the MLS is operationally useless. A signal that arrives six weeks early changes what a team can do.

Which market conditions affect signal reliability?

Signal behavior almost certainly varies by price tier, neighborhood density, local media environment, and macroeconomic conditions (particularly interest rate environments, which strongly affect move decisions). A model calibrated in a low-rate, high-appreciation market may behave differently in a rate-constrained market.

What geographic grain is actionable?

ZIP-code-level signals may be too coarse to be useful for a team that works at the neighborhood or street level. Whether the signals described here can be meaningfully disaggregated below the ZIP level — and whether the data volume supports that disaggregation — is an open question.

What does a false positive cost?

If a signal fires and no listing surge follows, the broker team may have allocated resources (mailers, door-knocking, agent time) to a geography that did not produce. Understanding the cost of false positives is necessary to set a useful signal threshold — a threshold that is too sensitive produces too many false positives; one that is too conservative misses real events.

Questions for broker teams reading this paper

The research questions above have analytical answers. The questions below have answers that only the broker teams themselves can provide — answers that are necessary before a pilot is designed.

1. What lead time — measured in weeks — would be sufficient to change how your team allocates resources across farm areas? Two weeks? Six weeks?
 2. In practice, when your team loses a listing to a competitor, what does that look like? Did the competitor arrive earlier, pitch differently, or know something about the seller's timeline that your team did not?
 3. Which of your current farm areas do you feel least informed about — where are you making the most assumptions with the least supporting data?
 4. What does your team currently use to assess market momentum in a neighborhood? How current is that information, and how confident are you in it?
 5. If a system produced a weekly "signal score" for each of your farm areas — a number indicating whether activity was likely to increase, decrease, or remain flat — how would your team actually use that information?
-

A closing note on intellectual honesty

Signal Engine's position going into pilot discussions is not that these signals work. It is that the hypothesis is worth testing rigorously in real markets alongside broker teams who have the domain knowledge to identify whether a correct prediction would have changed their behavior — and who have the MLS access to confirm whether predictions were actually correct. That is the only path to an honest answer.