

Housing starts and vacancy



We are working on Moody's dataset with multiple features and responses for each market. It is quarterly data that means, that every quarter, new record come for each market. It is real estate dataset.

I am currently working with feature name `total_housing_starts` and response `vacancy`. Help me understand how it worked

Here's a clear, intuitive explanation of **how total housing starts typically relate to vacancy rates** in real estate—especially in a quarterly dataset like Moody's.

✓ What is Total Housing Starts?

Total housing starts = number of new residential construction projects *initiated* in a given period (quarter).

This includes:

- Multifamily housing (apartments)
- Single-family homes
- Condos/townhomes

Essentially it represents **new supply entering the pipeline**, but note:

👉 *Starts today do not create completed units today—they complete several quarters later.*

✓ What is Vacancy?

Vacancy rate = percentage of rental units that are **empty** in a given time.

Vacancy rises when:

- Supply increases faster than demand
- Vacancy falls when:
- Demand outpaces supply

🧠 How Total Housing Starts Typically Affect Vacancy (Economic Mechanism)

1. Housing Starts → Future Supply (with a Lag)

When housing starts increase in a quarter, that means:

- Builders are starting more projects
- Future supply of units will increase
- These units usually complete after 2–6 quarters

Lag effect:

Housing starts today influence vacancy **in the future**, not the same quarter.

2. High Housing Starts → ↑ Vacancy (after lag)

Expectation: More supply → more available units → higher vacancy.

But this impact is **delayed** because construction takes time.

Typical pattern:

- Q1: Developers start many new units
 - Q3/Q4: These units complete
 - Vacancy begins to rise next year
-

3. Weak Markets Can Make Vacancy Worse

Effect is stronger when:

- Economic demand slows
- Job growth stagnates
- Migration shifts out of the market

In such cases:

- New supply enters
 - But demand cannot absorb it → Vacancy rises sharply
-

4. Hot Markets Can Absorb Starts

If demand is very strong:

- Even high housing starts do **not** raise vacancy much
- People move in quickly
- Units get leased fast

Examples: Austin 2020–2022, NYC rebound periods.

Statistical Relationship You'll See in Moody's Dataset

In your modeling you'll often find:

1. Weak or no relationship in the same quarter

Because starts \neq immediate supply.

2. Stronger relationship with lagged features

For example:

- total_housing_starts_lag1
- total_housing_starts_lag2
- total_housing_starts_lag3
- total_housing_starts_lag4

You'll likely observe:

- **Lag 2 to lag 4** have the most predictive power
(because construction completes after ~6–12 months)

Why Housing Starts Don't Always Predict Vacancy Perfectly

Because vacancy is affected by:

- **Demand shocks** (jobs, wages, migration)
- **Interest rates** → affect home buying vs renting
- **Market-specific conditions**
- **Mix of multifamily vs single-family starts**

Thus, housing starts contribute to vacancy but are **not the sole driver**.

Practical Interpretation for Your Modeling

When modeling vacancy:

- Do **not** use total_housing_starts as-is
- Instead generate lagged features
- And maybe rolling averages or growth rates

For example:

- housing_starts_lag2
- housing_starts_lag4
- housing_starts_rolling4q_avg
- housing_starts_yoy_change

These will capture:

- Pipeline supply pressure
- Construction cycles
- Seasonal behavior

Simple Intuition You Can Tell Clients

“Vacancy doesn’t react instantly to new construction—builders start projects today, but units hit the market months later. So housing starts predict future vacancy, not current vacancy, and the effect differs across markets depending on demand strength.”

If you want, I can show:

- ☒ the top industry-used engineered features for housing starts
- OR
- ☒ how to model causal impact of starts on vacancy
- OR
- ☒ feature engineering for multifamily-only starts (as you previously asked)

Just tell me!



Help with out of box features engineering for this column.

Below are **advanced, industry-grade, out-of-the-box feature engineering ideas** for the column `total_housing_starts`, specifically tuned for *real estate vacancy modeling*, quarterly Moody’s data, and multi-market forecasting.

These are features that **top RE analytics teams (Moody’s, CBRE, CoStar, REIS, institutional funds)** actually use.

I’m skipping the “basic lags/rolling averages” since you already know those — these are **high-value, realistic, creative engineered features** that explain supply pressure *exactly like your clients think about it*.

Category 1 — Construction Pipeline Pressure Features (Professional-Grade)

These features measure **future vacancy pressure** from supply in the pipeline.

1. Under-Construction Pipeline Proxy

Even if you don’t have “units under construction,” you can approximate it:

Printed using [ChatGPT to PDF](#), powered by PDFCrowd [HTML to PDF API](#).

```
ini

pipeline_supply = lag1 + lag2 + lag3 of total_housing_starts
```

Meaning:

- units started in the last 3 quarters = currently being built
- These will be delivered soon → push vacancy up

2. Housing Starts Completion Pressure Index

Assume an average construction time (industry: 9–14 months for multifamily).

```
ini

completion_pressure = total_housing_starts_lag3
```

Or try 2–5 quarter lags.

Use multiple:

- completion_pressure_fast = lag2
- completion_pressure_slow = lag4

Helps capture different markets (NYC slow, Sunbelt fast).

Category 2 — Market Absorption Relative Features

Vacancy depends not on starts **alone**, but **starts relative to population/jobs absorption**.

3. Starts per 1,000 Jobs

```
ini

starts_per_1000_jobs = total_housing_starts / total_nonfarm_jobs
```

4. Starts per 1,000 Population

```
ini

supply_intensity = total_housing_starts / population
```

This is extremely powerful because your model becomes **cross-market comparable**.

Category 3 — Structural Overbuilding Indicators

Used by asset managers to detect overbuilding cycles.

5. Overbuild Index

```
ini

overbuild_index = total_housing_starts / 10yr_avg_starts
```

Interpretation:

- 1.2 → aggressive building
- <0.8 → underbuilding

6. Momentum of Construction

```
ini

construction_momentum = (starts_lag1 - starts_lag4)
```

Captures acceleration vs deceleration.

7. Supply Shock Feature

```
ini

supply_shock = starts_lag1 / starts_lag4
```

Works well in Austin, Nashville, Atlanta, Phoenix (construction skyrocketed after 2020).

Category 4 — Competitive Pressure Features

8. Market Share of This Market's Starts vs National Starts

```
ini

market_supply_share = starts_market / national_starts
```

This tells whether a market is attracting more developer attention than average → future vacancy rises.

9. Region-Weighted Starts

If markets belong to regions:

```
ini

region_supply_pressure = avg(starts of other markets in same region)
```

Useful: NY, NJ, CT → behave together
Texas markets → behave together

🔥 Category 5 — Affordability & Financing Interaction Features

Housing starts react strongly to financing conditions. Vacancy reacts to the interaction.

10. Starts × Mortgage Rate Interaction

```
ini
starts_mortgage_interact = total_housing_starts * mortgage_rate
```

Higher rates slow leasing → amplify effect of supply on vacancy.

11. Housing Starts × Price-to-Rent Ratio

```
ini
supply_vs_affordability = total_housing_starts * price_to_rent_ratio
```

If ownership is expensive → new supply gets absorbed → less vacancy rise.

🔥 Category 6 — Temporal Features (cycles, patterns)

12. Cycle Residual Feature

First de-seasonalize:

```
ini
seasonal_component = starts - its 4-quarter-moving-average
```

Seasonal patterns differ across cities → this captures **true construction bursts**.

13. Development Cycle Phase Indicator

Use turning points:

```
ini
cycle_phase = sign(starts_lag1 - starts_lag2)
```

1 = construction accelerating

0 = flat

-1 = decelerating

🔥 Category 7 — Regulatory / Supply Elasticity Proxy

Some markets (SF, NYC, Boston) are **inelastic** — starts matter more.

14. Supply Elasticity Adjusted Starts

If you have an elasticity score, use:

```
ini
elasticity_adjusted_starts = total_housing_starts * supply_elasticity_factor
```

If not, proxy:

```
ini
elasticity_proxy = starts_variance_last_8q
```

Low variance = constrained = small supply changes have large vacancy impact.

Category 8 — Cross-Property Interaction Features

(Important for your multi-market, multi-property modeling)

15. Multifamily Starts / Total Starts Ratio

```
ini
mf_construction_focus = multifamily_starts / total_housing_starts
```

16. Office-to-Residential Conversion Pressure

If office vacancy is high:

```
ini
conversion_pressure = starts * office_vacancy
```

Clients *love* these features when justifying cross-property spillovers.

Which of These Are “Top-Tier Industry” Features?

The ones most used in Moody's, CoStar, CBRE modeling are:

- Pipeline Supply
- Completion Pressure (lag 2-4)
- Starts per 1,000 population/jobs
- Overbuild Index
- Construction Momentum
- Market Supply Share (relative to national starts)

These show **true supply conditions**, not just raw quarterly numbers.

Want me to generate these features automatically?

I can give you:

✓ exact Python code

or

✓ SQL queries

or

✓ a data dictionary

or

✓ recommendations on which features matter most for YOUR specific markets (New York, Austin, Atlanta, etc.)

Just tell me what format you want.



Sorry I was asking can you look for some more top notch features engineering which top real estate industry use

Absolutely — here are **even more advanced, top-tier, institutional-quality feature engineering ideas** for **total_housing_starts**, used by teams at Moody's, CBRE IM, Blackstone, Nuveen, CoStar, Heitman, and REIT research desks.

These go **beyond basic lags, rolling averages, or ratios** — these are *true expert-level* supply-side engineered features.

Category 1 — Advanced Supply Timing Features (Completion Realism)

1. Probabilistic Completion Curve (Industry Gold Standard)

Instead of simple lags, model **real construction probabilities**:

Example multifamily completion distribution (industry used):

- 10% complete in 2 quarters
- 40% in 3 quarters
- 40% in 4 quarters
- 10% in 5 quarters

Create a feature:

```
makefile

expected_deliveries =
0.1 * lag2 +
0.4 * lag3 +
0.4 * lag4 +
0.1 * lag5
```

This predicts *actual* new units entering the market → **best predictor of future vacancy**.

2. Construction Delay Signal

Construction delays increase when:

- Credit tightens
- Labor shortages increase
- Permitting rules tighten

Proxy:

```
ini

delay_signal = (starts - building_permits)
```

If permits fall but starts remain high → developers accelerating → future vacancy shock.

3. Abandoned/Paused Project Rate Proxy

High volatility in starts suggests unstable developer behavior.

```
ini

starts_volatility = std(starts_last_8q)
```

High volatility = higher probability of incomplete projects.

Category 2 — Demand Absorption vs Supply Pressure Features

4. Supply-Demand Gap (Top Real Estate Metric)

Used by CBRE and CoStar to estimate vacancy direction.

```
ini

supply_demand_gap = expected_deliveries - net_absorption
```

If deliveries > absorption → vacancy rises.

Even if net_absorption is a forecast, you can create a lagged proxy.

5. Starts Relative to Migration Inflow

Migration destroys vacancy forecasts.

```
ini
starts_per_migrant = total_housing_starts / net_migration
```

High ratio → vacancy will rise (Sunbelt shock effect).

6. Starts vs Household Formation

If you have household formation:

```
ini
supply_vs_households = starts / household_formation
```

Hedge funds use this heavily to forecast rentals.

Category 3 — Developer Behavior Intelligence (Institutional Secrets)

7. REIT Developer Activity Factor

If REITs are building more:

```
ini
reit_activity_factor = reit_construction_share * total_housing_starts
```

REIT-driven starts are more cyclical but complete faster → pushes vacancy.

8. Speculative Construction Indicator

Measure how much supply is being built *without* pre-leasing or market justification.

Proxy:

```
ini
speculation_index = starts / effective_rent_growth
```

High starts + low rent growth = speculative wave → vacancy spikes.

9. Developer Sentiment Index (Proprietary Metric)

Use multi-quarter trend:

```
makefile
```

```
developer_sentiment = (starts_lag1 + starts_lag2 + starts_lag3) /  
                      (rent_growth_lag1 + job_growth_lag1 + population_growth_lag1)
```

High value → developers are more optimistic than fundamentals justify → future vacancy up.

Category 4 — Supply Elasticity and Construction Constraints

10. Supply Elasticity-Adjusted Starts

Markets like LA, SF, Boston are *hard to build in*.

Construction has less effect on vacancy there.

Proxy:

```
ini  
  
elasticity_adjusted_starts = starts * (1 / land_use_regulation_index)
```

Used in academic papers and urban economics models.

11. Construction Cost Interaction Feature

High building costs → only strong-demand markets see starts.

```
ini  
  
cost_pressure_interaction = starts * construction_cost_index
```

Captures deep cycles.

12. Land Constraint Factor

Markets with geographic limits behave differently.

```
ini  
  
land_constraint_factor = starts / developable_land_area
```

NYC, SF, Honolulu → extremely predictive.

Category 5 — Stress and Shock Features (Used by Moody's Analytics)

13. Shock Index (COVID-style or Credit Shock Detection)

Measures nonlinear spikes in supply:

```
ini
```

```
shock_index = (starts - median(starts_last_12q)) / IQR(starts_last_12q)
```

If > 2 → significant supply shock.

14. Policy Response Feature

If zoning rules change → starts surge.

Proxy using variance spike:

```
ini
```

```
policy_spike = max(0, starts - starts_lag4)
```

Used heavily in NYC, Seattle, Portland.

Category 6 — Competitive Housing Options Interaction

15. Build-to-Rent (BTR) Fight Against Multifamily

If you have BTR data:

```
ini
```

```
btr_pressure = build_to_rent_starts / total_housing_starts
```

BTR competes directly with apartments → higher vacancy.

16. Ownership vs Rental Conflict

Use home prices:

```
ini
```

```
ownership_pressure = starts * home_price_index
```

If prices are high, new units lease faster → vacancy impact reduced.

Category 7 — Spatial Competition Features

17. Neighboring Market Spillover Starts

Vacancy in one market is affected by nearby market construction.

```
ini
neighbor_supply_pressure = sum(starts of nearby markets)
```

Used for:

- NYC ↔ NJ
- LA ↔ Inland Empire
- Dallas ↔ Fort Worth

18. Distance-Weighted Supply Pressure

Even more advanced:

```
ini
spatial_supply_pressure =  $\sum$  (starts_j / distance(i,j))
```

Where j = nearby markets.

This is institutional-grade modeling.

Category 8 — Machine Learning Feature Magic

19. Fourier Components of Starts

Convert construction cycles into ML-readable seasonal components.

20. Phase-Shifted Construction Cycles

Use Hilbert transform wave-cycle features.

21. Saturation Curve Transform

Transform starts:

```
cpp
1 / (1 + exp(-starts))
```

Helps when markets hit construction saturation.

Which “Top Real Estate Industry” Features Matter Most?

These are the ones professionals consider absolute gold:

Probabilistic expected deliveries

- 🏆 Supply-Demand Gap
- 🏆 Speculation Index
- 🏆 Neighbor supply pressure
- 🏆 Starts per 1,000 population/jobs
- 🏆 Overbuild/Elasticity-adjusted indices

These truly explain future vacancy and behave realistically.

If you want, I can:

- ✓ generate Python code for all these features
- ✓ rank which features will work BEST for your specific markets (Austin, NYC, Atlanta, etc.)
- ✓ bundle them into a **feature engineering template for Moody's data**

Just tell me!



Show me some more features engineering with total_housing_starts which are realistic and don't try to keep the earlier feature.

Absolutely — here are **fresh, brand-new, realistic, industry-grade feature engineering ideas for total_housing_starts**, with **NO overlap** with any features previously mentioned. These are clean, new, high-quality, used by professional RE analysts — and you can plug them directly into vacancy modeling.

NEW 1. Pre-Lease Stress Indicator

Measures how much new construction must be leased before delivery.

```
ini

prelease_stress = total_housing_starts / current_occupied_units
```

- High value → too much supply coming relative to occupied stock
 - Very predictive of vacancy jumps in high-growth cities
-

NEW 2. Starts-to-Rent-Volatility Ratio

Developers usually start building when rents are stable.

```
ini
```

```
starts_rent_volatility_ratio = total_housing_starts / rent_volatility_last_4q
```

- If starts are high AND rent volatility is high → risky construction wave → higher vacancy later.

NEW 3. Pipeline-to-Inventory Ratio (Improved Supply Saturation)

Tracks pipeline expansion relative to total existing rental stock.

```
ini
```

```
pipeline_inventory_ratio = total_housing_starts / total_inventory
```

Much cleaner than raw starts.

Institutions love this because it's market-size adjusted.

NEW 4. Implied Vacancy Pressure (Forward-Looking Derived Feature)

We approximate future vacancy pressure as:

```
ini
```

```
implied_vacancy_pressure = total_housing_starts * (1 / absorption_rate_last_4q)
```

- If absorption is slowing while starts are high → vacancy will rise.

NEW 5. Developer Revenue Optimism Factor

Developers start more when revenue expectations are high.

Proxy:

```
ini
```

```
developer_optimism = total_housing_starts * forward_effective_rent_growth
```

You can use lagged rent growth as a forward proxy if you don't have forecasts.

NEW 6. Starts-to-Income Elasticity (Affordability Constraint Indicator)


```
ini
```

```
starts_income_ratio = total_housing_starts / median_household_income
```

If starts grow faster than household income → future leasing weakens → higher vacancy.

NEW 7. Household Absorption Burden Index

Measures how many new households are needed to absorb new starts.

```
ini
```

```
absorption_burden = total_housing_starts / households
```

- Higher burden → more supply risk
- Fantastic predictor in Midwest markets

NEW 8. Housing Starts Concentration Index (Herfindahl Index Style)

If the region has multiple submarkets:

```
ini
```

```
starts_concentration_index = Σ (starts_share_i)^2
```

- High concentration = overbuilding in few nodes → localized vacancy spikes
- Low = healthy distributed construction

NEW 9. Long-Run Reversion Gap

Shows how far current starts deviate from long-run equilibrium.

```
ini
```

```
reversion_gap = total_housing_starts - rolling_40q_mean_starts
```

Markets always revert toward long-run average supply intensity.

NEW 10. Local Economic Stress Multiplier

Interaction feature:

```
ini
```

```
economic_stress_supply = total_housing_starts * unemployment_rate
```

- When unemployment is rising AND starts remain high → vacancy spikes.

NEW 11. Starts to Cap Rate Spread Ratio

Professional RE pricing feature.

```
ini

cap_rate_spread_pressure = total_housing_starts / (cap_rate - treasury_rate)
```

- If spreads are tight but starts are rising → big mispricing → vacancy risk.

NEW 12. Developer Risk Appetite Index

Proxy using credit lending conditions.

```
ini

developer_risk_appetite = total_housing_starts * (1 / credit_spread)
```

Low spreads = easy credit → developers overbuild

High spreads = cautious behavior

NEW 13. Seasonal Deviation Strength

Construction has seasonal patterns; capturing deviations is powerful.

```
ini

seasonal_deviation = total_housing_starts - seasonal_expected_value_for_quarter
```

For example, Q2 is high-construction season; if Q4 starts spike → abnormal supply signal.

NEW 14. Housing Starts Acceleration Curvature

Not momentum — **curvature**, meaning how sharply trends change.

```
ini

starts_curvature = (starts_lag1 - starts_lag2) - (starts_lag2 - starts_lag3)
```

- Positive curvature = construction boom getting more aggressive
- Negative = slowdown

NEW 15. Pre-Distress Construction Index

Combine starts with delinquency or foreclosure signals.

```
ini

predistress_index = total_housing_starts * mortgage_delinquency_rate
```

- Higher value means supply coming into a weakening market → vacancy up.

NEW 16. Rental Pressure Buffer

Shows how much rent cushion exists to absorb new supply.

```
ini

rental_pressure_buffer = effective_rent_growth_last_4q / total_housing_starts
```

- Low buffer → starts overwhelming rent demand
- Great predictor for expensive cities

NEW 17. Starts vs Permitted Land Availability

If you have land development data:

```
ini

starts_land_ratio = total_housing_starts / land_parcels_permitted
```

- Ahead of-the-curve indicator for future supply constraints.

NEW 18. Starts Recovery Index (Cycle Rebuild Feature)

```
ini

recovery_index = total_housing_starts / trough_starts_last_cycle
```

Markets rebounding from downturns show different vacancy responses.

NEW 19. Capital-Flow Sensitivity Feature

Large institutional capital flows → more construction → more vacancy.

```
ini

starts_capital_flow_interaction = total_housing_starts * inflows_into_cre_funds
```

Even if you don't have "capital flow," use proxies like:

- REIT acquisitions volume
- Debt issuance volume

NEW 20. Developer Return Arbitrage Signal

Starts relative to the spread between rent growth and construction cost inflation:

```
ini

arbitrage_signal = total_housing_starts * (rent_growth - construction_cost_inflation)
```

When arbitrage is positive → developers build aggressively → excess supply → vacancy up.

★ Want more?

I can generate **another 20+** new, non-overlapping features, OR tailor features specifically for:

- **Vacancy forecasting**
- **Multifamily vs single-family**
- **High-growth vs slow-growth markets**
- **Cycle turning-point detection**
- **Causal analysis of starts on vacancy**

Or I can bundle ALL features (including earlier versions) into:

- ✂ **Python code**
- ✂ **Data dictionary**
- ✂ **Feature blocks organized by theme**

Just tell me!