

Understand HPI and vacancy rate



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 `Home_price_Index_all_transactions` and response `vacancy_rate`. Help me understand how it works

Perfect 👍 let's unpack this step by step — since you're working with **Moody's real estate dataset**, quarterly, and focusing on:

- **Feature:** `Home_price_Index_all_transactions`
- **Response:** `vacancy_rate`

Let's carefully understand **what this feature represents, how it relates to vacancy_rate, and how to extract insights or engineer features** from it.

🧩 1. What is `Home_price_Index_all_transactions` (HPI)?

The **Home Price Index (HPI)** measures the **price level of residential properties** over time in a given market.

- **"All transactions"** means it includes both:
 - *repeat sales* (homes sold multiple times),
 - *first-time sales* (new properties sold first time).

So it's a **broad indicator** of how home prices are trending — **upward (appreciation)** or **downward (depreciation)** — across the market.

🏠 2. What is `vacancy_rate`?

`vacancy_rate` measures the **percentage of available rental properties (or units)** that are **vacant** (not occupied) during a given quarter.

It's a strong **demand indicator** — if demand is weak, vacancy rises; if demand is strong, vacancy falls.

🔄 3. The Relationship Between HPI and Vacancy Rate

Economic Phase	HPI Trend	Vacancy Rate Trend	Intuition
Boom / Growth Period	▲ Increases	▼ Decreases	Home prices rise as demand for buying and renting increases. Fewer vacant properties.
Downturn / Recession	▼ Decreases	▲ Increases	Lower demand → home prices fall → higher vacancy.

Economic Phase	HPI Trend	Vacancy Rate Trend	Intuition
Overheated Market	▲ Rapid increase	▲ May also increase	Sometimes prices grow too fast → affordability drops → some properties stay unsold or unrented.
Stabilizing Market	↔ Flat or slow increase	↔ Stable	Equilibrium between supply and demand.

So the relationship can be:

- **Negative correlation** during normal cycles (HPI ↑ → Vacancy ↓)
- **Nonlinear** during shocks or bubbles.

⚙️ 4. Analytical Ways to Study This

Here are ways you can analyze or model the relationship:

a. Correlation Analysis

Check correlation between `Home_price_Index_all_transactions` and `vacancy_rate` per market:

```
python

df.groupby('market')[['Home_price_Index_all_transactions', 'vacancy_rate']].corr()
```

You'll likely see a **negative correlation** in most stable markets.

b. Lag Relationship

Vacancy might not respond immediately to price changes — there's often a **lag** (1-2 quarters).

You can test this:

```
python

df['HPI_lag1'] = df.groupby('market')['Home_price_Index_all_transactions'].shift(1)
```

Then check how `HPI_lag1` predicts `vacancy_rate`.

c. Growth Rates (Feature Engineering)

Raw HPI values don't mean much across markets — growth or change does.

Compute quarterly or annual growth:

```
python

df['HPI_qoq_change'] = df.groupby('market')['Home_price_Index_all_transactions'].pct_change()
df['HPI_yoy_change'] = df.groupby('market')['Home_price_Index_all_transactions'].pct_change(4)
```

These growth features are very informative for **predicting vacancy trends**.

d. Interaction with Other Variables

Combine HPI with other market indicators:

- `HPI_to_income_ratio`
- `HPI_growth × interest_rate`
- `HPI_growth × unemployment_rate`

These can capture **affordability** or **economic stress** effects that drive vacancy up or down.

5. Example Interpretations

Let's say:

- In **Austin**, HPI increased by 5% QoQ while vacancy fell from 7% to 5%.
→ Indicates strong demand and limited supply.
- In **New York**, HPI surged 12% but vacancy also rose from 4% to 6%.
→ Indicates prices may have outpaced demand; affordability issue emerging.

6. Summary

Concept	Meaning
<code>Home_price_Index_all_transactions</code>	Reflects property value trends in a market
<code>vacancy_rate</code>	Reflects demand/supply balance of occupied vs. vacant units
Typical Relation	Inverse ($\text{HPI} \uparrow \rightarrow \text{Vacancy} \downarrow$)
Analysis Tips	Look at lags, percentage change, and interactions with macro variables
Feature Ideas	<code>HPI_qoq_change</code> , <code>HPI_yoy_change</code> , <code>HPI_lag1</code> , <code>HPI_growth × income</code>

If you'd like, I can now help you **design top-tier feature engineering ideas** specifically from `Home_price_Index_all_transactions` — similar to what institutional real estate analysts or Moody's internal econometric teams would do.

Would you like me to show you that next?



Help with out of box features engineering for this column.

Perfect 🍌 — since you're working on **real estate forecasting** with Moody's data and focusing on `Home_price_Index_all_transactions` (HPI) as your feature, let's go beyond the usual "lag" or "percent change" type features.

Below I've grouped **out-of-the-box (industry-grade)** feature engineering ideas used by **Moody's, CBRE, CoStar, and REIT analysts** to capture the deeper dynamics between **HPI** and **vacancy_rate**.

1. Growth-based features (price momentum)

Feature	Formula / Description	Intuition
Quarter-over-Quarter % Change	$(\text{HPI}_t - \text{HPI}_{\{t-1\}}) / \text{HPI}_{\{t-1\}}$	Measures short-term price momentum.
Year-over-Year % Change	$(\text{HPI}_t - \text{HPI}_{\{t-4\}}) / \text{HPI}_{\{t-4\}}$	Captures medium-term appreciation trend.
3-Quarter Rolling Growth	Average of last 3 quarters' QoQ growth	Smooths volatility.
Price Acceleration	Difference between current and previous growth rates	Detects turning points in market momentum.

Use these to signal whether the market is *heating up* (sustained positive growth) or *cooling*.

2. Lag & Lead Features (temporal dynamics)

Feature	Description	Why it matters
HPI_lag1, HPI_lag2, HPI_lag4	Previous 1, 2, 4 quarters' HPI values	Vacancy often reacts with delay.
$\Delta\text{HPI}_{\text{future}}$ (lead)	HPI change 1-2 quarters ahead	Can test for predictive relationships using Granger causality.

Helps model "how long it takes" for price changes to reflect in vacancy.

3. Volatility & Stability Features

Feature	Formula / Method	Use case
Rolling Std Dev (σ)	Std of HPI growth over 4-8 quarters	Captures market uncertainty.
Coefficient of Variation (CV)	$\sigma / \text{mean}(\text{HPI})$	Standardized volatility indicator.
Rolling Range (max-min)	Range of HPI in past year	Detects cyclical price swings.

Markets with higher volatility often have higher vacancy volatility too.

4. Affordability & Stress Metrics

Combine HPI with **income**, **rent**, or **interest rates** to measure affordability pressure:

Feature	Formula / Description	Intuition
Price-to-Income Ratio	$\text{HPI} / \text{Median_Household_Income}$	High ratio → affordability issues → rising vacancy.
Price-to-Rent Ratio	$\text{HPI} / \text{Rent_Index}$	If buying is expensive vs renting → demand shifts.
Interest-Rate Adjusted Price Growth	$\text{HPI_growth} - \text{Mortgage_Rate_change}$	Measures net cost pressure on ownership.

💡 These composite ratios are widely used in institutional property models.

🌐 5. Market Comparison & Spatial Features

Feature	Description	Insight
Relative HPI to National Mean	$\text{HPI_market} / \text{HPI_national}$	Indicates local over/undervaluation.
Cross-Market Spread	Difference between market's HPI growth and neighboring markets' average	Detects migration or substitution effects.
Rank of HPI Growth among all Markets	Market's percentile rank each quarter	Useful for identifying outperformers vs laggards.

Example: "Austin HPI 20% above national → may attract investors but risk of oversupply."

🔄 6. Cyclicalty & Decomposition Features

Feature	Method	Purpose
Seasonal Decomposition (Trend + Cycle)	STL decomposition	Separates long-term vs cyclical effects.
Fourier Terms	sin/cos waves to model seasonality	Captures repeating cycles.
Turning Point Flags	Detect local maxima/minima using slope changes	Early warning for market reversal.

🔍 Especially helpful since real estate cycles often run 5–8 years.

🧩 7. Elasticity & Sensitivity Features

Use first differences to measure **price elasticity** relative to macro factors:

Feature	Formula	Measures
Elasticity_to_Unemployment	$\Delta \text{HPI} / \Delta \text{Unemployment}$	How sensitive are home prices to job market?
Elasticity_to_Income	$\Delta \text{HPI} / \Delta \text{Income}$	Captures income responsiveness.

Real estate investment teams love these for stress testing.

📊 8. Cumulative or Index-based Features

Feature	Formula	Meaning
Cumulative Growth Since Baseline	$(\text{HPI_t} / \text{HPI_baseline_2010}) - 1$	Long-term appreciation.
Z-score of HPI	$(\text{HPI_t} - \text{mean}(\text{HPI_market})) / \text{std}(\text{HPI_market})$	Measures deviation from typical level.
Rolling CAGR (Compound Annual Growth Rate)	Over last 1–3 years	Smooth long-term momentum.

9. Structural Change Features


Feature	Description
Breakpoint Dummy	1 if structural break detected in HPI trend (Chow test, Bai-Perron)
Post-Crisis Indicator	Separate periods (pre/post COVID, GFC, etc.)
Slope Shift Variable	Captures permanent trend change after major events.

These are powerful for macro-driven models.

10. Interaction Features with Vacancy Drivers

Combine HPI with variables like:

- `HPI_growth × business_bankruptcies`
- `HPI_growth × construction_activity`
- `HPI_growth × population_growth`
- `HPI_lag1 × interest_rate_lag1`

 These show how price movements interact with supply/demand stressors to affect vacancy.

Putting It All Together (Example Code)

Here's a compact Python snippet to generate advanced HPI features:

```
python

df = df.sort_values(['market', 'quarter'])

df['HPI_qoq'] = df.groupby('market')['Home_price_Index_all_transactions'].pct_change()
df['HPI_yoy'] = df.groupby('market')['Home_price_Index_all_transactions'].pct_change(4)
df['HPI_volatility'] = df.groupby('market')['HPI_qoq'].rolling(4).std().reset_index(level=0, drop=True)
df['HPI_acceleration'] = df.groupby('market')['HPI_qoq'].diff()
df['HPI_zscore'] = (df['Home_price_Index_all_transactions'] -
                    df.groupby('market')['Home_price_Index_all_transactions'].transform('mean'))
/ \
                    df.groupby('market')['Home_price_Index_all_transactions'].transform('std')
df['HPI_relative_national'] = df['Home_price_Index_all_transactions'] / \
                    df.groupby('quarter')
['Home_price_Index_all_transactions'].transform('mean')
```

Would you like me to show **which of these features are most predictive for vacancy_rate**, along with reasoning for **which ones institutional models (like Moody's CRE forecast systems) prioritize**?



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

Ah! Now I understand — you want the **premium, industry-grade, cutting-edge feature engineering** that **top real estate analysts and firms like Moody's, CBRE, CoStar, and REIT modeling teams** actually use when working with **Home Price Index** to predict vacancy, rent growth, or investment risk. Let's go deep.

I'll organize this by **feature types**, including some that are rarely discussed outside professional real estate analytics.

1 Advanced Temporal Features

1. Dynamic Lag Features

- Not just `lag1` or `lag4`.
- Compute **weighted moving lags** that decay over time:

$$HPI_{weighted_lag} = \sum_{i=1}^n HPI_{t-i} \cdot w_i, \quad w_i = 0.9^{i-1}$$

- Captures **recent trends more heavily**, used in CRE (commercial real estate) forecasting.

2. Rolling Momentum / Trend Indicators

- `HPI_MA_short` = 3-quarter rolling mean
- `HPI_MA_long` = 12-quarter rolling mean
- Feature: **MA ratio** = `HPI_MA_short` / `HPI_MA_long`
- High ratio → upward momentum; low → market cooling.

3. Quarterly Slope / Trend Change

- Fit a linear regression of HPI over the past N quarters per market.
- Use slope as feature: rising slope → prices accelerating.

4. Volatility Breakout Flags

- Identify **periods when HPI changes** > 2σ of recent 4 quarters.
- Flags speculative markets or stress periods.

2 Market Comparisons & Relativity Features

1. HPI vs National / Peer Markets

- `HPI_relative` = `HPI_market` / `HPI_national`
- `HPI_spread` = `HPI_market` - `mean(HPI_peers)`

2. Rank Features

- Rank each market's HPI growth percentile within all markets.
- Captures *relative performance*: very useful for investment screening.

3. Distance-Weighted HPI

- Take nearby markets' HPI trends weighted by geographic or economic distance.
- Captures spillover effects (e.g., growth in Austin affecting San Antonio).

3 Cyclicalty & Seasonality Features

1. STL Decomposition

- Split HPI into trend, seasonal, residual.
- Use residuals as a feature — captures **unexpected shocks**.

2. Fourier Series Terms

- Encode quarterly seasonality: $\sin(2\pi t/4)$, $\cos(2\pi t/4)$
- Helps ML models capture cyclical price swings.

3. Turning Points

- Detect local maxima/minima in HPI trends.
- Feature: **1 if market at local peak, -1 if trough**.
- Highly predictive for vacancy spikes.

4 Stress / Affordability Features

1. Price-to-Income Ratio

$$HPI / Median_Income$$

- Directly used by Moody's and REIT teams. High ratio → rising vacancy risk.

2. Price-to-Rent Ratio

$$HPI / Rent_Index$$

- Detects market overpricing relative to rental yields.

3. Mortgage-Adjusted HPI

$$HPI_{adjusted} = HPI \times (1 + MortgageRate_change)$$

- Shows impact of borrowing cost changes on housing demand.

4. Elasticity Features

- How HPI growth responds to macro variables: unemployment, income, interest rates.
- Features like $HPI_growth / \Delta Unemployment$ or $HPI_growth / \Delta Income$.

5 Structural & Regime Change Features

1. Breakpoints & Trend Shifts

- Detect using **Bai-Perron** or **Chow test**.
- Feature: 1 if post-breakpoint, 0 otherwise.

2. Crisis / Event Dummies

- COVID-19, 2008 Financial Crisis, interest rate shocks.
- Features can capture abnormal market conditions.

3. Slope Shift Indicators

- Feature: change in slope of HPI growth over previous 4–8 quarters.
 - Early warning for overheating or cooling markets.
-

6 Interactions / Composite Features

1. HPI × Construction Activity

- High HPI + high construction → future vacancy risk.

2. HPI × Business Bankruptcies / Job Loss

- Rising prices + economic stress → predicts rising vacancy.

3. HPI Growth × Population Growth / Migration

- Fast-growing population can absorb HPI rise → moderates vacancy.
-

7 Advanced Statistical Features

1. Rolling Z-Score

$$Z = \frac{HPI_t - \mu_{rolling}}{\sigma_{rolling}}$$

- Detects over- or undervaluation relative to historical trend.

2. CAGR (Compound Annual Growth Rate)

- Over 1–3 years to capture long-term trends.

3. Rolling Skew / Kurtosis

- Detect markets with unusual HPI distribution → predicts instability.
-

8 Real Estate AI / ML Tricks Used by Top Firms

- **Lagged Interaction Features:** combine multiple lags of HPI with unemployment, income, or interest rates.
- **Cumulative Price Pressure:** sum of HPI growth over last 4–8 quarters.

- **Volatility Regimes:** classify market into **stable**, **volatile**, **speculative** based on rolling std.
- **Relative Momentum Signals:** compare short-term vs long-term trend to capture *trend reversals*.

♦ Summary of “Top-Notch” Features for HPI

Category	Examples
Temporal	Weighted lags, momentum ratio, acceleration, slope
Market Comparison	Relative HPI, spread vs peers, rank percentile
Cyclicity	STL residual, Fourier terms, turning points
Affordability	Price-to-Income, Price-to-Rent, Mortgage-adjusted HPI
Stress / Regime	Breakpoint flags, crisis dummies, slope shifts
Interactions	HPI × construction, HPI × unemployment, HPI × population
Statistical / ML	Rolling Z-score, CAGR, skew/kurtosis, volatility regime

If you want, I can **create a concrete Python template** that generates **all these top-tier features** from your `Home_price_Index_all_transactions` column — ready to plug into your vacancy prediction model. This would be exactly how **industry analysts engineer features**.

Do you want me to do that?



So me some more features engineering with `Home_price_Index_all_transactions`.

Absolutely! Let’s go **even deeper** into advanced feature engineering for `Home_price_Index_all_transactions` (HPI) — the kind **top real estate analysts and Moody’s CRE teams** would use. I’ll break it into **categories with concrete feature ideas** beyond standard lags and growth rates.

1 Advanced Temporal Dynamics

1. Exponential Weighted Moving Average (EWMA)

```
ini
HPI_EWMA = α * HPI_t + (1-α) * HPI_EWMA_{t-1}
```

- Captures **smoothed trend with more weight on recent quarters**.

2. Price Acceleration

```
ini
HPI_acceleration = (HPI_t - HPI_{t-1}) - (HPI_{t-1} - HPI_{t-2})
```

- Measures **speed of price change**, which can signal bubbles or cooling periods.

3. Rolling Momentum Indicator

```
ini

HPI_momentum_4q = HPI_t - HPI_{t-4}
```

- 1-year price momentum.

4. Rolling Slope

- Fit a linear regression over past 4–8 quarters; slope is a feature.
- Positive slope → rising market, negative → falling.

2 Volatility & Uncertainty Metrics

1. Rolling Std Dev

```
ini

HPI_vol_4q = std(HPI_t-3..HPI_t)
```

- Captures **market uncertainty**.

2. Coefficient of Variation

```
ini

HPI_cv = std(HPI_rolling) / mean(HPI_rolling)
```

- Normalized volatility; compares markets of different scales.

3. Rolling Max/Min Range

```
ini

HPI_range_4q = max(HPI_rolling) - min(HPI_rolling)
```

- Detects **cyclical swings** or speculative pressure.

3 Relative & Comparative Features

1. HPI vs National Average

```
ini

HPI_relative = HPI_market / HPI_national
```

- Shows **market over/underperformance**.

2. HPI Spread vs Peer Markets

```
ini

HPI_spread = HPI_market - mean(HPI_similar_markets)
```

- Detects **divergence from comparable regions**.

3. Percentile / Rank

```
ini

HPI_rank = percentile rank of HPI growth among all markets
```

- Highlights **top/bottom performing markets**.

4 Affordability & Stress Features

1. Price-to-Income Ratio

```
nginx

HPI / Median_Household_Income
```

- Higher ratio → affordability stress → vacancy risk.

2. Price-to-Rent Ratio

```
nginx

HPI / Rent_Index
```

- High → overpricing vs rental demand → potential vacancy increase.

3. Mortgage-Adjusted Price

```
ini

HPI_adjusted = HPI * (1 + ΔMortgageRate)
```

- Reflects **financing cost pressure** on buyers.

5 Cyclicalty & Seasonality

1. STL Decomposition Residual

- Use residual component from HPI trend-seasonal decomposition to capture **unexpected shocks**.

2. Fourier Terms

```
scss

sin(2πt/4), cos(2πt/4)
```

- Captures quarterly cyclical in HPI.

3. Turning Point Flags

- 1 if HPI at local peak, -1 if local trough.

6 Structural & Regime Change

1. Breakpoint Flags

- Identify structural breaks using **Bai-Perron or Chow test**.
- Feature: 1 after structural change.

2. Crisis/Event Dummy

- Pre/post major market event (COVID, 2008 crisis).

3. Slope Shift Feature

- Change in trend slope over past 4–8 quarters.

7 Interaction & Composite Features

1. HPI × Vacancy Lag

```
nginx
HPI_growth × Vacancy_lag1
```

- Measures how **price growth interacts with existing vacancy pressure**.

2. HPI × Construction / Inventory

- High growth + high construction → future oversupply risk.

3. HPI × Economic Stress

- HPI_growth × Unemployment Rate
- Captures **price vs local economy tension**.

4. Population / Migration Weighted HPI

- HPI × Population_Growth
- Rising HPI in growing markets may have **lower vacancy risk**.

8 Statistical / AI-inspired Features

1. Rolling Z-Score

```
ini
Z = (HPI_t - mean_rolling(HPI)) / std_rolling(HPI)
```

- Overpriced vs underpriced relative to recent history.

2. CAGR over 1–3 years

```
ini
```

$$\text{CAGR} = (\text{HPI}_t / \text{HPI}_{\{t-4n\}})^{(1/n)} - 1$$

- Long-term growth rate.

3. Rolling Skew & Kurtosis

- Detect **non-normality in HPI distribution**, often predictive of market stress.

✓ These are **all features actually used by top analysts** for:

- Predicting vacancy rates
 - Forecasting rent / price growth
 - Risk analysis for CRE portfolios
-

If you want, I can **write a Python template** that will **compute all these advanced HPI features automatically for your quarterly dataset** — fully ready for ML models or vacancy prediction.

Do you want me to do that next?