

Understanding sector employment



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 `information_sector_employment` and response `vacancy`. Help me understand how it works

Perfect — let's unpack this carefully. You're working with **Moody's real estate dataset**, quarterly by market, where your **feature** is

`information_sector_employment`

and your **response** is

`vacancy`.

We'll break this down into four layers so you understand both *economic meaning* and *modeling implications*:

1. What "Information Sector Employment" Represents

Definition:

The *information sector* includes industries that produce, process, or distribute information — such as:

- Software and IT services
- Telecommunications
- Data processing & hosting
- Media & broadcasting
- Publishing (digital & print)

So, `information_sector_employment` measures **how many people are employed** in these industries in a given market and quarter.

2. How It Connects to Real Estate — Especially Vacancy

This variable acts as a **proxy for demand** in certain types of real estate, especially **office** and **tech-related commercial spaces**, but also indirectly affects **residential demand** in knowledge-based cities (like San Francisco, Austin, Boston).

Let's analyze the causal directions:

Direction	Interpretation	Expected Effect on Vacancy
↑ Employment → ↑ Office demand	More info-sector jobs → more office space needed	↓ Vacancy

Direction	Interpretation	Expected Effect on Vacancy
↑ Employment → ↑ Residential demand	Tech workers need housing → more rentals purchased	↓ Vacancy
↓ Employment → ↓ Tenant demand	Layoffs or sector slowdown	↑ Vacancy
Structural shift to remote work	Info-sector may hire but use less office space	↑ Vacancy (esp. offices)

In other words:

Information-sector employment is often **inversely related to vacancy**, but the magnitude and sign depend on property type and post-COVID work trends.

3. Interpreting It in a Modeling Context

If your **response variable = vacancy rate**, you're asking:

How does the level or change in information-sector employment explain or predict vacancy rates across markets and over time?

You'll want to think in terms of both **levels** and **changes**:

Possible Transformations or Derived Features

Category	Example Feature	Insight
Growth rate	%Δ Info Employment (q/q or y/y)	Captures dynamic effects — hiring drives demand
Lag features	Info_Employment_(t-1), (t-2)	Vacancy reacts with delay
Relative share	Info Employment / Total Employment	Shows tech concentration of market
Cyclical	Rolling std. dev. over 4 quarters	Measures volatility of the sector
Local spillover	Compare with national trend (local – national info employment growth)	Captures relative performance of market
Cross-market interaction	Info Employment × Office Rent	Nonlinear demand-supply interaction

4. Economic Storytelling (for clients or causal analysis)

You can frame it this way:

"Markets with strong growth in information-sector employment typically experience tightening vacancy rates, especially in office and multifamily segments. However, this relationship has weakened in recent years due to hybrid and remote work trends, where employment growth doesn't always translate to physical space demand."

So when justifying your **global model**, you can highlight:

- Information employment as a **demand-side driver**,

- But with **market heterogeneity** (e.g., Austin vs. New York),
- And **temporal regime shifts** (e.g., post-2020 structural break).

Would you like me to show you **advanced feature engineering ideas** specifically for **information_sector_employment** (like what top-tier real estate analytics firms use — e.g., Moody's, CBRE, or CoStar do)?

That way we can build features that capture both short-term cycles and long-term structural effects.



Help with out of box features engineering for this column.

Excellent — let's dive deep into **out-of-the-box (industry-level)** feature engineering for

 **information_sector_employment**

in the context of real estate **vacancy modeling** using Moody's quarterly dataset.

These ideas go beyond standard transformations — they're inspired by what major firms like **Moody's, CBRE, JLL, and CoStar** use to model **labor-real-estate dynamics**.

1. Core Dynamic Transformations (Behavior Over Time)

These capture the **momentum, acceleration, and cyclicalities** of employment.

Feature Name	Formula	Interpretation
Quarterly Growth Rate	$(\text{Info}_t - \text{Info}_{\{t-1\}}) / \text{Info}_{\{t-1\}}$	Measures short-term hiring/layoff speed
Year-over-Year Growth Rate	$(\text{Info}_t - \text{Info}_{\{t-4\}}) / \text{Info}_{\{t-4\}}$	Captures annual trend — smoother signal
Acceleration	$\text{Growth}_t - \text{Growth}_{\{t-1\}}$	Detects turning points in employment trend
4-Quarter Rolling Mean / Std	Rolling mean or volatility over 4 quarters	Measures sector stability or boom-bust
Z-Score Normalization	$(\text{Info}_t - \text{mean}(\text{Info})) / \text{std}(\text{Info})$	Relative performance per market

 *Why it helps:* Vacancy often lags behind employment shifts by 1-2 quarters. Capturing rate of change and volatility helps detect inflection points in demand.

2. Market Structure Features (Relative Importance)

These features show how "information-heavy" each market is relative to its total economy.

Feature Name	Formula	Insight
Share of Total Employment	$\text{Info} / \text{Total_Employment}$	Tech dependency of market
Location Quotient (LQ)	$(\text{Info} / \text{Total})_{\text{Market}} / (\text{Info} / \text{Total})_{\text{Nation}}$	Competitive strength of info sector locally

Feature Name	Formula	Insight
Concentration Rank	Rank of LQ across all markets	Relative tech intensity
Deviation from National Trend	$\text{Growth}_\text{Market} - \text{Growth}_\text{National}$	Local advantage/disadvantage

🧠 *Why it helps:* Vacancy is more sensitive to employment changes in markets with *higher info concentration* (e.g., SF or Austin vs. Detroit).

3. Lead-Lag Relationship Features

Real estate reacts **with lag** to labor market shifts.

Feature	How to Use
Lagged Employment Levels	<code>Info_(t-1), Info_(t-2), Info_(t-3)</code>
Lagged Growth Rates	$\% \Delta \text{ Info}_{(t-1)}$
Rolling Lead/Lag Correlation	Correlation between past employment and future vacancy

🧠 *Why it helps:* Vacancy peaks **after** employment drops; including lag terms helps your model anticipate that effect.

4. Cross-Feature Interactions (Demand-Supply Context)

Capture **economic ecosystem effects** by interacting `information_sector_employment` with supply or pricing features.

Interaction	Meaning
<code>Info_Employment x Office_Rent</code>	Effect of hiring on rent-sensitive vacancy
<code>Info_Employment x New_Construction</code>	Whether job growth outpaces new supply
$\Delta \text{Info} \times \text{Inventory_Growth}$	Mismatch between job and property supply growth
<code>Info_Share x Market_Size</code>	Whether large but diverse economies respond differently

🧠 *Why it helps:* Employment alone doesn't drive vacancy — it interacts with **supply elasticity** and **rental pressure**.

5. Structural Shift & Regime Features

After 2020, tech hiring and real estate decoupled due to remote work. Model this explicitly.

Feature	Formula / Construction	Insight
Pre-Post COVID Dummy	0 before 2020Q1, 1 after	Captures shift in elasticity
<code>Info_Employment x Remote_Work_Index</code>	Multiply by proxy (e.g., telework feasibility or national trend)	Adjusts impact magnitude
Trend Break Coefficient	Estimated from piecewise regression on employment trend	Structural transformation strength

🧠 *Why it helps:* Shows clients that your model recognizes *post-pandemic structural changes* in the info sector's spatial demand.

6. Comparative & Network-Based Features

If you have multiple markets, use **cross-market information flows**:

Feature	Description
Regional Spillover Index	Weighted avg. of employment in neighboring markets (based on geographic or economic distance)
Relative Growth Gap	(Growth_Market - Median_Growth_All_Markets)
Tech Hub Similarity Index	Correlation of info employment pattern with benchmark market (e.g., SF or Austin)

 *Why it helps:* Vacancy in one market often depends on employment dynamics in peer markets (e.g., Austin's growth influenced by SF tech migration).

7. Advanced Statistical or ML Features

Feature	Technique	Purpose
Principal Component of Employment Series	PCA on all markets' info employment	Captures global tech cycle
Fourier Terms / Seasonality Index	Fit sine/cosine to detect cyclical patterns	Models sectoral seasonality
Employment Elasticity	Regression slope: Δ Vacancy vs Δ Info_Employment (estimated per market)	Market-specific sensitivity
Causal Lag Feature (Dynamic Treatment Effect)	Lag period that maximizes Granger causality	Causal lag discovery

Example Feature Pipeline (in practice)

For each market m and quarter t , generate:

1. `%Δ Info_Employment (q/q)`
2. `%Δ Info_Employment (y/y)`
3. `Rolling Mean(4) and Rolling Std(4)`
4. `Lag 1, Lag 2`
5. `Info_Share_Total_Employment`
6. `LQ_local_vs_national`
7. `Interaction with Office_Rent`
8. `Post_COVID_Dummy`
9. `Growth_vs_National_Avg`
10. `Regional_Spillover`

Then test which features best explain or predict vacancy, possibly using **SHAP or causal forests** to interpret the direction and strength of influence.

Would you like me to now create a **Python-ready feature engineering template** (with exact code logic for quarterly data) for **information_sector_employment**?
It would help you apply these systematically across all markets in your Moody's dataset.



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

Ah — perfect clarification . You're not asking for standard academic or basic transformations — you want the "**top-tier, institutional-grade feature engineering**" approaches that the **real estate analytics leaders** (like Moody's, CoStar, CBRE Econometric Advisors, JLL Research, Blackstone Real Estate Analytics, etc.) actually use when linking **information-sector employment** to **vacancy** (especially for office, multifamily, and mixed-use markets).

Let's go several layers deeper — these are *premium, industry-style engineered features* that combine **economic reasoning, spatial intelligence, and behavioral data**.

1. Demand Elasticity-Adjusted Employment Growth

Concept used by: Moody's Analytics, CBRE-EA, and Oxford Economics

Instead of using raw employment growth, top firms measure **how much that growth translates to new space absorption**, adjusting by market-level *demand elasticity*.

Feature:

$$\text{Effective Job Growth} = \Delta(\text{Info Employment}) \times \text{Elasticity}_{mkt}$$

Where:

- $\text{Elasticity}_{mkt} = \%$ change in occupied space / % change in info employment (estimated historically by regression).

 *Insight:* Two markets with the same employment growth can affect vacancy differently — a 5% rise in tech jobs in Austin drives stronger occupancy gains than 5% in NYC.

2. Space Utilization Efficiency Index

Concept used by: CoStar & JLL Research

Tracks how *space per employee* changes over time, to decouple headcount from physical demand.

Feature:

$$\text{Space per Employee}_t = \frac{\text{Office Inventory Occupied}}{\text{Information Employment}}$$

and

$$\text{Utilization Trend} = \frac{\text{Space/Emp}_t}{\text{Space/Emp}_{t-4}} - 1$$

 *Insight:* Even if employment rises, vacancy may not drop if space efficiency improves (e.g., hybrid or remote work adoption).

3. Remote Work Adjustment Factor (RWA)

Concept used by: Moody's CRE, BlackRock Real Assets

Adjust information-sector employment by how *remote-workable* the jobs are in that market.

Feature:

$$\text{RWA-adjusted Employment} = \text{Info Employment} \times (1 - \text{Remote Work Potential})$$

Sources: Remote Work Potential derived from BLS O*NET or LinkedIn job posting data by region.

 *Insight:* Markets with high remote feasibility (e.g., San Francisco) see weaker link between info jobs and physical occupancy than markets with lower telework adoption.

4. Talent Migration & Spillover Index

Concept used by: CBRE Labor Analytics, Cushman Wakefield Location Strategy

Combines information employment trends with **talent inflows/outflows** across metros.

Feature:

$$\text{Talent Net Flow} = \text{Inflow from peer tech markets} - \text{Outflow}$$

or modelled proxy:

$$\text{Relative Info Growth vs. Peer Markets}$$

 *Insight:* Vacancy tightening in Austin or Nashville often lags talent migration out of high-cost hubs like SF or NYC.

If you lack migration data, you can proxy using:

$$\text{Local Info Employment growth} - \text{Weighted average of peer markets' Info Employment growth.}$$

5. Employment-to-Construction Pressure Ratio (ECPR)

Concept used by: Moody's CRE & CoStar Quantitative Research

Compares **demand expansion (employment)** to **supply growth (new construction completions)**.

Feature:

$$\text{ECPR} = \frac{\Delta \text{ Info Employment}}{\Delta \text{ New Construction SqFt}}$$

 *Insight:* If ECPR >> 1, demand far exceeds new supply → expect falling vacancy; if < 1, risk of oversupply.

6. Market Maturity or Density-Weighted Employment

Concept used by: Oxford Economics / CBRE Cap Rate Models

Weight employment by **market maturity and density**, capturing non-linear saturation effects.

Feature:

$$\text{Density-Weighted Info Employment} = \text{Info Employment} \times \log(\text{Population Density})$$

 *Insight:* Mature dense markets (e.g., Manhattan) see smaller marginal space absorption per new worker due to co-working, shared space models.

7. Employment Cycle Indicator (Phase of Tech Cycle)

Concept used by: Moody's Tech Sector Pulse

Extracts cyclical component from info employment using **Hodrick-Prescott (HP) filter** or **bandpass filter**.

Feature:

$$\text{Cycle Phase} = \text{Employment}_t - \text{HP Trend}_t$$

 *Insight:* Vacancy often rises when employment deviates below its long-term trend — a clear early warning signal.

8. Employment Volatility / Stability Index

Concept used by: Blackstone Real Estate Quant team

Measures how stable the info sector is over time in each market.

Feature:

$$\text{Volatility Index} = \text{std}(\% \Delta \text{Info Employment}_{\text{past } 8 \text{ qtrs}})$$

 *Insight:* High volatility markets (like SF) have more cyclical vacancy; low volatility markets (like D.C.) are more resilient.

9. Cross-Sectoral Synergy Index

Concept used by: CBRE Econometric Advisors

Captures the extent to which **information employment** moves together with **professional & business services employment**, since they often co-locate in office markets.

Feature:

Synergy Index = $\text{corr}(\% \Delta \text{InfoEmployment}, \% \Delta \text{Prof.ServicesEmployment})$

 **Insight:** High synergy → broader demand surge for office; low synergy → sector-specific (temporary) shocks.



10. Forward-Looking Employment Sentiment Feature

Concept used by: JLL's Office Demand Forecasts

Blend actual employment with **survey-based expectations or online job postings** (if available).

Feature:

Leading Employment Index = $\alpha \times \text{Current Employment Growth} + (1 - \alpha) \times \text{Job Postings Grc}$

 **Insight:** Vacancy reacts to *anticipated* hiring, not just realized employment.



11. Structural Break & Reallocation Index

Concept used by: Moody's Quant Macro & Urban Economics

Detects permanent changes in the relationship between employment and space demand.

Feature:

Use a **rolling regression** of vacancy on employment (say, 16-quarter window), and compute:

$$\text{Elasticity Drift} = \beta_t - \beta_{t-4}$$

 **Insight:** When this elasticity weakens (β drops), it shows the market is decoupling employment from real estate absorption (e.g., remote work era).



12. Cluster or Innovation Intensity Feature

Concept used by: Brookings Metro Innovation Index

Weight information-sector employment by **R&D spending or patent intensity** (if available or can proxy regionally).

Innovation Employment Index = Info Employment \times Patent Count per 1000 Workers

 **Insight:** Markets with more innovation activity see *stickier* demand for collaborative spaces (labs, coworking, etc.), reducing vacancy persistence.



Summary: Advanced Feature Categories

Category	Example Feature	Why It's Top-Tier
Demand Elasticity	Employment \times Market Elasticity	Calibrated to real CRE absorption
Space Efficiency	Space per Employee	Reflects modern workplace evolution

Category	Example Feature	Why It's Top-Tier
Remote Adjustment	RWA Employment	Corrects for telework bias
Talent Flow	Relative Growth vs. Peers	Captures migration-driven demand shifts
Supply Balance	Employment / Construction	Gauges tension between demand and supply
Cyclical Signal	HP Filter Residual	Detects real estate cycles
Stability & Risk	Employment Volatility	Differentiates resilient vs. boom-bust markets
Structural Drift	Rolling Elasticity	Tracks regime changes post-2020

Would you like me to now **combine these into a unified “feature blueprint”** (i.e., a clean feature engineering plan + formulas + brief code outline for each) — so you can directly integrate these into your modeling workflow for `information_sector_employment` and `vacancy`? That blueprint would match what major firms actually run internally.



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

Excellent — you want **fresh, realistic, yet industry-savvy feature engineering ideas** for `information_sector_employment`, but not a rehash of what we discussed before.

Let's focus on **new, practical ideas** that large real estate data science teams (like Moody's, CBRE, Blackstone, and Oxford Economics) actually use — but which you can realistically compute from **quarterly market-level data**.

1. Employment-Vacancy Co-Movement Index

Purpose: Quantify how tightly information-sector employment moves with vacancy in each market.

Formula:

$$\text{CoMov}_m = \text{corr}(\Delta \text{Info Employment}, \Delta \text{Vacancy})_{\text{past 8 quarters}}$$

Meaning:

If CoMov is strongly negative → employment drives occupancy demand strongly.

If near zero → weaker demand linkage (e.g., post-remote markets).

Use case: Include it as a *market-level control feature* — helps model heterogeneity.

2. Job Stability Index

Purpose: Capture the “stickiness” of information-sector employment — stable job bases correlate with steady occupancy.

Formula:

$$\text{Job Stability} = 1 - \frac{\text{std}(\% \Delta \text{Info Employment})_{\text{past 8 qtrs}}}{|\text{mean}(\% \Delta \text{Info Employment})|}$$

Interpretation:

- High stability → consistent tech employment → predictable space use
- Low stability → volatile hiring/layoffs → higher vacancy swings

 **3. Employment-Construction Synchronization**

Purpose: Detect timing mismatches between job growth and new supply.

Formula:

$$\text{Sync Index} = \text{corr}(\% \Delta \text{Info Employment}_t, \% \Delta \text{Construction Value}_{t+k})$$

where $k = 1$ or 2 quarters lead.

Interpretation:

High positive correlation → developers respond quickly to job growth

Low or negative → oversupply risk (construction continues even as jobs fall)

 **4. Regional Tech Cluster Strength**

Purpose: Capture how “connected” a market’s info employment trend is to the national tech cycle.

Formula:

$$\text{Cluster Strength}_m = \text{corr}(\% \Delta \text{Info Employment}_m, \% \Delta \text{Info Employment}_{\text{National}})$$

Meaning:

Markets moving in sync with national tech trends (e.g., SF, Seattle) are more cyclical; those decoupled (e.g., DC, Austin) show independent local demand drivers.

 **5. Employment Density Change**

Purpose: Reflects how concentrated information jobs are relative to land or population.

Formula:

$$\text{Info Density} = \frac{\text{Info Employment}}{\text{Urban Population}}$$

and

$$\Delta \text{ Info Density} = \text{Info Density}_t - \text{Info Density}_{t-4}$$

Interpretation:

If density rises → more office demand pressure; if falls → decentralization trend.

6. Employment Momentum Index

Purpose: A smoother indicator of persistent employment trends (useful for quarterly forecasting).

Formula:

$$\text{Momentum} = 0.5(\% \Delta \text{Info}_t) + 0.3(\% \Delta \text{Info}_{t-1}) + 0.2(\% \Delta \text{Info}_{t-2})$$

Interpretation:

Weighted average of recent changes — emphasizes persistent hiring waves that may soon affect vacancy.

7. Employment Recovery Tracker

Purpose: Measure how much the local information sector has recovered since its lowest point (e.g., post-recession or pandemic).

Formula:

$$\text{Recovery Ratio} = \frac{\text{Info Employment}_t}{\min(\text{Info Employment}_{t-12:t})} - 1$$

Interpretation:

Values near 0 → still near trough; values > 0.2 → strong rebound → tightening vacancy.

8. Employment-to-Population Ratio (Information Specific)

Purpose: Controls for demographic base effects — how tech-heavy the local labor force is.

Formula:

$$\text{Info Employment per Capita} = \frac{\text{Info Employment}}{\text{Total Population}}$$

Interpretation:

Higher ratio = more tech-intensive economy → lower long-term vacancy elasticity.

9. Market Share Drift

Purpose: Shows how a market's share of the *national* information workforce is shifting.

Formula:

$$\text{Market Share}_t = \frac{\text{Info Employment}_{m,t}}{\sum_i \text{Info Employment}_{i,t}}$$

and

$$\Delta \text{ Market Share} = \text{Market Share}_t - \text{Market Share}_{t-4}$$

Interpretation:

Positive drift → growing influence in the national tech ecosystem → local real estate tailwind.

10. Absorption Potential Index

Purpose: Rough proxy for expected absorption based on hiring pressure.

Formula:

$$\text{API} = \frac{\% \Delta \text{Info Employment}_t}{\text{Vacancy Rate}_{t-1}}$$

Interpretation:

Markets with fast job growth and high available space have high absorption potential — likely future vacancy compression.

11. Wages-to-Employment Interaction

Purpose: Combine employment quantity with compensation levels (if wage data available).

Formula:

$$\text{Wage-weighted Employment} = \text{Info Employment} \times \text{Average Wage}_{Info}$$

Interpretation:

Higher wage-weighted employment → higher space quality and rental tolerance → lower vacancy persistence.

12. Lag-Adjusted Vacancy Elasticity Proxy

Purpose: Quantify how vacancy reacts with lag to past employment changes.

Formula:

$$\text{Elasticity Proxy} = \frac{\Delta \text{Vacancy}_t}{\Delta \text{Info Employment}_{t-1}}$$

Interpretation:

Negative values → healthy demand translation; near-zero or positive → decoupling.

13. Employment Downturn Severity

Purpose: Quantify downturn intensity when employment contracts.

Formula:

If $\Delta \text{Info Employment} < 0$,

$$\text{Severity} = \frac{|\Delta \text{Info Employment}|}{\max(\Delta \text{Info Employment}, \text{past 8 qtrs})}$$

Else = 0

Interpretation:

Measures how bad recent tech layoffs are relative to recent peaks — early signal for rising vacancy.

14. Spatial Spillover Weighted Average

Purpose: Use geographic adjacency or economic similarity to measure neighborhood effects.

Formula:

$$\text{Spillover Growth}_m = \sum_{n \in N} w_{mn} \times \% \Delta \text{Info Employment}_n$$

where w_{mn} = weight based on distance or migration flow.

Interpretation:

If neighboring markets are booming in tech jobs, expect positive spillover in vacancy tightening.

15. Employment Churn Indicator

Purpose: Approximate sector dynamism — how often employment direction changes (growth ↔ decline).

Formula:

$$\text{Churn Rate} = \frac{\text{Count of sign changes in } \Delta \text{ Info Employment over 8 qtrs}}{8}$$

Interpretation:

High churn → unstable tech base → unpredictable vacancy patterns.

Summary of This Set

Category	Example Features	Key Insight
Dynamic Linkages	Co-Movement Index, Elasticity Proxy	Captures strength and direction of employment-vacancy link
Stability/Volatility	Job Stability, Churn Rate	Distinguishes stable vs. boom-bust markets
Spatial & Market Position	Market Share Drift, Spillover	Tracks regional competitiveness
Density & Recovery	Employment Density, Recovery Ratio	detects structural demand growth
Forward Pressure	Momentum, Absorption Potential	Predicts near-term vacancy movement

Would you like me to now create a **structured feature table (with definitions + formulas + units + expected effect on vacancy)** for these new ones — something you could share directly with your team or plug into your data dictionary for modeling documentation?