

# Unemployment and vacancy rates



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 unemployment\_rate and response vacancy\_rate.

Help me understand how it works

Perfect — this is a great place to start. Let's break down the relationship between **unemployment rate** and **vacancy rate** in the context of a **real estate dataset** that updates **quarterly** across **multiple markets** (like Moody's Analytics data).

We'll go step by step so that you get both the *economic intuition* and the *statistical modeling perspective*



## 1. The Structure of Your Data

Since it's **quarterly panel data**, you likely have something like this:

Market	Quarter	Unemployment_Rate	Vacancy_Rate	Other_Features...
Atlanta	2020Q1	5.0	10.2	...
Atlanta	2020Q2	8.2	12.5	...
Austin	2020Q1	4.1	9.5	...
Austin	2020Q2	6.8	11.0	...
...	...	...	...	...

Each row is a **snapshot of a market in a specific quarter** — meaning your model must capture **both time variation** (economic cycles, recovery, etc.) and **cross-sectional variation** (differences between markets).

## 2. Understanding the Two Variables

### A. Unemployment Rate (feature / predictor)

- Represents the **percentage of the labor force that is unemployed** in that market and quarter.
- It is a **macroeconomic indicator** that reflects the **health of the local economy**.
- Higher unemployment usually means weaker job markets → lower income → less spending and business expansion → reduced demand for real estate.

## B. Vacancy Rate (response / dependent variable)

- Percentage of **available real estate units (apartments, offices, retail spaces, etc.)** that are **vacant** at the end of the quarter.
- High vacancy rates mean **lower demand** or **excess supply**.
- It's a **key performance indicator** for real estate health — especially for investors and lenders.

## 3. The Intuitive Economic Relationship

Generally, there's a **positive relationship** between unemployment and vacancy rate:

- When **unemployment increases**, companies cut jobs, shrink office space, and consumers spend less.  
→ This leads to **higher vacancy rates**.
- When **unemployment decreases**, the economy grows, companies expand, people move for jobs.  
→ This leads to **lower vacancy rates**.

So, if we visualize:

```
nginx
```

```
Unemployment ↑ → Vacancy Rate ↑
Unemployment ↓ → Vacancy Rate ↓
```

However, this relationship may **not be immediate** — sometimes **lag effects** exist.

## 4. The Role of Time (Lagged Effects)

Vacancy doesn't instantly respond to unemployment changes.

For example:

- If unemployment spikes in Q1, office leases might not be dropped until Q2 or Q3.
- Hence, **vacancy\_rate\_t** might depend on **unemployment\_rate\_(t-1)** or even **(t-2)**.

👉 You can model this with **lagged features**, e.g.:

```
cpp
```

```
vacancy_rate_t = α + β1 * unemployment_rate_t + β2 * unemployment_rate_(t-1) + ε
```

This captures **both contemporaneous and delayed effects**.

## 5. Modeling Perspective

You can model this relationship using various approaches depending on your goal:

Goal	Model Type	Example Formula
Simple correlation	OLS regression	<b>vacancy_rate ~ unemployment_rate</b>

Goal	Model Type	Example Formula
Time and market effects	Fixed effects panel model	<code>vacancy_rate ~ unemployment_rate + C(market) + C(quarter)</code>
Forecasting	Time series model	<code>vacancy_rate_t = f(unemployment_rate_t, unemployment_rate_(t-1), ...)</code>
Nonlinear patterns	ML model (e.g., LightGBM, EBM)	Flexible relationships

Example (Python-statsmodels syntax):

```
python

model = smf.ols("vacancy_rate ~ unemployment_rate + C(market) + C(quarter)", data=df).fit()
```

This controls for:

- Market-level heterogeneity (some cities always have higher vacancy),
- Quarter-level shocks (e.g., COVID-19 effect).

## 6. Interpretation of the Coefficient

Suppose your regression gives:

```
ini

vacancy_rate = 1.5 + 0.8 * unemployment_rate
```

- $\beta = 0.8$  means:

For every 1% increase in unemployment, the vacancy rate increases by 0.8 percentage points (on average).

This is your **marginal effect** — and it quantifies the relationship strength.

## 7. Market-Specific Differences

Each market may react differently:

- Tech-heavy markets (like Austin) might recover faster.
- Industrial-heavy markets (like Detroit) might respond more slowly.

You can model that using **interaction terms**:

```
python

vacancy_rate ~ unemployment_rate * C(market)
```

This allows the slope ( $\beta$ ) to vary across markets.

## 8. What You Can Explore Next

1. **Check correlations** between unemployment and vacancy rate per market.
2. **Add lag features** to capture delayed effects.
3. **Control for macro variables** (e.g., GDP growth, inflation, construction permits).
4. **Use visualization** — plot both rates over time per market.
5. **Test causality** using methods like Granger causality or causal inference (your domain 😊).

If you like, I can show you:

- 📊 how to visualize the dynamic relationship between these two variables for a given market
- or 📈 how to statistically estimate their relationship using Python (e.g., fixed effects regression with lags).

Which one would you like to do next?



Help with out of box features engineering for this column.

Excellent — that's exactly where real insight starts 🔍

Since you're working with **quarterly real estate data** and focusing on **unemployment\_rate** as a feature to predict **vacancy\_rate**, we can go *beyond the raw number* of unemployment and engineer **richer, more expressive features**.

Let's explore this in layers — from straightforward transformations to *economically inspired* and *market-dynamic* features that can make your model far smarter and more interpretable.

## 1. Basic Transformations (Direct Statistical Enrichment)

Feature Name	Formula / Idea	Rationale
<code>unemployment_rate_lag1</code>	Previous quarter's unemployment	Real estate reacts with delay. Captures lagged effects.
<code>unemployment_rate_lag2</code>	Two-quarter lag	Captures longer adjustment cycles.
<code>unemployment_rate_diff</code>	<code>unemployment_rate - unemployment_rate_lag1</code>	Measures short-term <i>momentum</i> (is unemployment rising or falling?).
<code>unemployment_rate_pct_change</code>	<code>(unemployment_rate - unemployment_rate_lag1) / unemployment_rate_lag1</code>	Scales the change relative to previous level — interpretable as a growth rate.
<code>unemployment_rate_ma4</code>	4-quarter moving average	Smooths cyclical noise, captures the trend.

Feature Name	Formula / Idea	Rationale
unemployment_rate_std4	Rolling 4-quarter std dev	Captures volatility — how unstable the local job market is.

👉 These are classic, simple, and **effective time-based enrichments**.

## 🧠 2. Economic Cycle Features (Macro-Context Engineering)

Feature	Description	Economic Meaning
unemployment_above_national	$\text{local\_unemployment} - \text{national\_unemployment}$	Captures how each market performs vs. the nation — useful because vacancy is sensitive to <i>relative performance</i> .
unemployment_gap_to_trend	$\text{unemployment\_rate} - \text{rolling\_trend}$ (e.g., HP filter or moving average)	Deviation from long-term trend → measures “cyclical pressure” on vacancies.
recession_flag	1 if $\text{unemployment\_rate} > \text{threshold}$ (e.g., 7%)	Marks downturn periods → vacancy dynamics often shift in regime.
job_market_recovery_speed	Slope of unemployment over last 4 quarters	Captures rate of improvement or deterioration of labor market.

These features reflect **economic health trajectories**, not just raw values.

## 🏙️ 3. Market-Relative Dynamics

Each market behaves differently; hence, **relative change** features are gold.

Feature	Formula / Description
market_unemployment_zscore	Standardize within each market (subtract mean, divide by std). Helps model learn deviations from market norm.
market_rank_in_unemployment	Rank each market by unemployment each quarter (1 = lowest, N = highest). Relative competitiveness of local economy.
market_unemployment_delta_from_median	$\text{unemployment\_rate} - \text{median\_unemployment\_rate\_all\_markets}$

→ These features let your model **compare markets within the same time slice**, improving interpretability across geographies.

## ⌚ 4. Time & Seasonality Aware Features

Quarterly real estate data often exhibits **seasonal patterns** (e.g., hiring surges, relocation cycles).

Feature	Description
quarter_sin / quarter_cos	Encode seasonality using sine/cosine of quarter number (1–4).
year_on_year_change	Compare Qx vs same quarter last year: $(U_t - U_{\{t-4\}})$ .
rolling_min_unemployment_4q	Minimum unemployment over past year → captures economic “bottom” cycles.

Feature	Description
<code>time_since_min_unemployment</code>	Number of quarters since last lowest unemployment → gauges recovery age.

These help models capture **cyclic and recovery patterns**.

## 5. Cross-Feature Interactions (Causal / Behavioral Enrichment)

Now, combine unemployment with **other macro or real estate variables** to form interactions that reflect deeper mechanisms.

Feature	Formula / Example	Insight
<code>unemployment_x_gdp_growth</code>	Product of unemployment & GDP growth	Captures effect of job loss in expanding vs. contracting economy.
<code>unemployment_x_construction_activity</code>	Interaction with new supply	High unemployment + high new supply → vacancy skyrockets.
<code>unemployment_change_x_rent_growth</code>	Change in unemployment interacting with rent trend	Indicates affordability stress and tenant movement.

## 6. Derived Stability Indicators (Volatility / Stress)

Real estate markets are slow-moving but **react strongly to instability**.

Feature	Description
<code>unemployment_volatility_8q</code>	Rolling std over 8 quarters (2 years).
<code>unemployment_shock_flag</code>	1 if
<code>shock_intensity</code>	Magnitude of deviation from 4-quarter average.

These detect **stress or shocks**, which often precede sudden jumps in vacancy.

## 7. Advanced (Optional) — Latent Dynamics & Embeddings

If you want to push beyond traditional engineering:

- Use **Principal Component Analysis (PCA)** across macro variables (including unemployment, GDP, CPI, etc.) to extract “macro factors.”
- Or, train a small **temporal autoencoder** on unemployment time series to extract latent trend features per market.
- You can also fit an **ARIMA** or **Kalman filter** to unemployment and use the **predicted next quarter unemployment** as an anticipatory feature.

## 8. Example Implementation in Python (Pandas)

python

```
import pandas as pd

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

# Basic lags and changes
df['unemployment_rate_lag1'] = df.groupby('market')['unemployment_rate'].shift(1)
df['unemployment_rate_lag2'] = df.groupby('market')['unemployment_rate'].shift(2)
df['unemployment_rate_diff'] = df['unemployment_rate'] - df['unemployment_rate_lag1']
df['unemployment_rate_pct_change'] = df['unemployment_rate_diff'] / df['unemployment_rate_lag1']

# Rolling stats
df['unemployment_rate_ma4'] = df.groupby('market')['unemployment_rate'].rolling(4).mean().reset_index(level=0, drop=True)
df['unemployment_rate_std4'] = df.groupby('market')['unemployment_rate'].rolling(4).std().reset_index(level=0, drop=True)



# Relative to national
national_unemp = df.groupby('quarter')['unemployment_rate'].mean()
df['unemployment_above_national'] = df['unemployment_rate'] - df['quarter'].map(national_unemp)
```

## 9. Best Practice Tip

When you engineer many features:

- Keep them **economically interpretable** (helps trust with clients and executives).
- Watch for **multicollinearity** (many features will be correlated).
- Use **feature importance** tools (e.g., SHAP, permutation importance) to validate which ones matter.


If you want, I can:


1.  Generate a **ready-to-use Python feature-engineering pipeline** for unemployment\_rate (with all above logic),
2.  Or help you **visualize how engineered features behave** across markets (e.g., lag vs. current correlation).


Which one would you like to explore next?



Yes.

Perfect  — before I build it for you, could you please confirm **which one** of these you'd like me to start with?

Option A →  **Python feature-engineering pipeline** for unemployment\_rate (with lag, trend, volatility, national-relative, etc.) — something you can directly plug into your real estate model.

Option B →  **Visualization notebook** — showing how engineered unemployment features behave across markets (e.g., how lag or volatility links to vacancy\_rate).

Once you choose, I'll generate the complete, ready-to-run code and explain how to interpret the results.



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

Great question — “top-notch” real estate industry feature engineering tends to pull in lots of external data, latent signals, spatial/graph structure, and forward-looking indicators. Here are some advanced feature engineering ideas that leading firms / researchers use, plus some concrete suggestions you might adapt for your unemployment ↔ vacancy\_rate setting.

## Advanced Feature Engineering Ideas in Real Estate

These go beyond standard lag, moving averages, etc. They tap into spatial, structural, alternative data, market regime effects, etc.

### 1. Spatial / Geographic Features & Proximity Derived Measures

- **Distance to market center / CBD:** how far a property or market region is from the central business district or economic core. Vacancy tends to behave differently by distance.
- **Access to infrastructure / transport nodes:** proximity to highways, public transit, major roads, rail stations.
- **Walkability / transit score:** aggregating amenities within radius around market centroid (or around properties) — shops, schools, hospitals, parks.
- **Neighborhood quality / crime rates / environmental risk:** crime stats, air pollution, flooding zones.
- **Spatial spillover / neighboring market stats:** vacancy / unemployment in adjacent or nearby markets as features (spatial lag). e.g., if surrounding markets have high vacancy or unemployment, might spill in.

### 2. Supply / Stock Dynamics

Vacancy doesn't just respond to unemployment; supply side matters a lot.

- **New construction completions** (quarterly new stock) / permits issued in last x quarters.
- **Absorption rate:** how fast existing units are getting occupied.
- **Age distribution of stock:** old vs new properties may have different vacancy behaviors.
- **Renovation / turnover rate:** how often units are vacated or renovated.

### 3. Market Sentiment / Forward-looking Signals

These help you anticipate shifts before they fully show in standard metrics.

- **Leading indicators:** e.g., job postings, consumer confidence, business sentiment surveys.
- **Interest rates / mortgage rate spreads:** cost of financing affects demand, thus vacancy.
- **Credit availability / lending conditions:** tighter credit means slower real estate investment or purchase, pushes up vacancies.



- **Building permits** as mentioned above, but particularly forward-looking permits (projects under construction) reflect expected future supply.

#### 4. Regime / Policy / Events Features

Real estate responds differently under different regimes — booms, busts, interest rate cycles, pandemics, etc.

- **Recession / downturn flags:** e.g., flag when unemployment, GDP, etc cross thresholds.
- **Policy changes:** rent control changes, zoning law shifts, tax incentives.
- **Macro shocks:** COVID or other economic disruptions — maybe include indicator for shock events.

#### 5. Alternative Data, Remote Sensing, Big Data

These are being used more and more in top real estate analytics.

- **Night-time lights / satellite imagery:** proxy for economic activity in regions. [MDPI](#)
- **Foot traffic / mobility data:** smartphone mobility or location data to see how many people are around shopping districts, business centers.
- **Remote sensing of building density / land usage:** how built-up vs green vs vacant land around.

#### 6. Clustering / Market Typologies

Not all markets are the same; grouping helps capture non-linear, heterogeneous effects.

- **Cluster markets based on demographic & economic structure:** e.g., high tech vs manufacturing vs service-heavy markets.
- **Cluster property types:** residential vs commercial vs mixed; small units vs large units. Vacancies behave differently.
- **Use unsupervised methods** (K-means, hierarchical, embeddings) to derive “market archetypes”.

#### 7. Non-linear and Interaction Features

Capturing interactions often improves real estate models significantly.

- **Interaction between unemployment and supply:** e.g.,  $unemployment\_rate * new\_construction\_rate$ .
- **Interaction between unemployment and demographic trends:** e.g., youth unemployment rate vs total unemployment; population growth rate  $\times$  unemployment.
- **Threshold effects / piecewise features:** e.g., unemployment above certain level  $\Rightarrow$  different slope.

#### 8. Temporal / Lagged / Trend Features (but more nuanced)

Some “best practice” lags and trends.

- **Lagged features at multiple horizons** (1, 2, 4 quarters) not just simple recent lag.
- **Rolling growth rates over different windows** (e.g., 4q vs 8q).
- **Rolling up-/down-trend indicators:** is unemployment trending up/down? Use slope over past n quarters.

- **Seasonality adjustments:** especially for vacancy (some property types may have seasonal variation). For example, seasonal decomposition of the vacancy rate and using seasonal residuals.

## 9. Synthetic / Composite Indicators

Instead of single raw variables, build composite indices which may better capture underlying latent phenomena.

- **Affordability index:** combining rent levels, local incomes, mortgage rates.
- **Demand-supply ratio:** effective demand index / supply completions.
- **Labor market tightness:** ratio of job vacancies to unemployment (if data available).
- **Real wage growth:** wage growth adjusted for inflation; helps understand capacity to pay rent.

## 10. Error / Residual Features

Once you build a baseline model, sometimes the **residuals** carry interesting info useful for predicting further changes.

- Residuals from a simple trend model of vacancy might indicate “undervalued” or “over-tight” markets.
- Use residuals of unemployment forecast vs actual — the surprise in unemployment may influence vacancy.

## Concrete Suggestions: Features You Could Add Based on unemployment\_rate + Your Real Estate Vacancy Model

Given your dataset has unemployment\_rate and vacancy\_rate per market per quarter, here are specific engineered features using those plus external/additional data you may pull in:

- **Unemployment Forecast Error:** build a simple model to forecast unemployment for market, and include the error (surprise). Vacancy might respond more to surprises than to levels.
- **Unemployment spread:**  $\text{market\_unemployment\_rate} - \text{national\_unemployment\_rate}$  or maybe region/university or state level if your markets are nested.
- **Supply Pressure / Permit Lag:** For each market, compute number of new units permitted or expected to be completed in next 1-2 quarters, normalized by existing stock. Combine with unemployment: e.g.,  $\text{supply\_pressure} = \text{new\_units} / \text{total\_stock}$ ; then  $\text{interaction\_feature} = \text{unemployment\_rate} * \text{supply\_pressure}$ .
- **Economic Momentum:** Rate of change of unemployment over multiple lags (e.g., quarter on quarter, year over year). Maybe include second derivative (i.e., acceleration of unemployment).
- **Housing Affordability / Income Ratios:** If you have median household income in each market, compute ratio of unemployment to median income, or the ratio of rent (or rent growth) to income.
- **Relative Vacancy Gap:** how does a market's vacancy compare with its historical mean / trend + standard deviation. E.g.,  $\text{vacancy\_gap} = \text{vacancy\_rate} - \text{vacancy\_trend\_4q}$ ,  $\text{vacancy\_stddev\_4q}$ .
- **Unemployment Volatility Clusters:** Mark markets in high volatility periods (where unemployment fluctuates more). Maybe cluster markets by their unemployment volatility.

- **Interaction: Demographics × Unemployment:** If you can get data on population growth rate, age distribution, migration (in/out), then interaction terms.
- **Lagged Vacancy Responses:** In addition to unemployment lag, include lagged vacancy itself (vacancy\_rate\_(t-1), etc.), maybe a “vacancy momentum” feature.
- **Shock Flags:** indicator variables for big jumps in unemployment (above a threshold) — maybe these amplify vacancy effects.
- **Macro Factor Features:** inflation, interest rate changes — how cheap or expensive borrowing is.
- **Spatial Spillover:** vacancy / unemployment in nearby markets (if you can define “nearby”, e.g., geographical distance or economic similarity), as spillover features.

If you like, I can put together a **set of 10-15 such “industry-grade” engineered features with formulas/code** tailored to your data, so you can test them. Do you want me to generate that list for your unemployment ↔ vacancy model?



Sure.

Excellent 🍌 — let’s build you a **top-tier feature engineering pack** that mirrors what major real-estate analytics groups (like Moody’s, CoStar, CBRE Econometrics, and REIS) typically use.

We’ll focus on creating **interpretable + data-science-ready** features that link **unemployment\_rate** → **vacancy\_rate** across **markets and quarters**.

## A. Temporal & Momentum Features

Capturing dynamics, shocks, and cyclical phases.

Feature Name	Formula / Logic	Meaning
unemp_lag1, unemp_lag2, unemp_lag4	Lag unemployment by 1, 2, 4 quarters	Real estate reacts slowly; captures delayed impact
unemp_qoq_change	<code>unemployment_rate - unemp_lag1</code>	Quarter-over-quarter labor market momentum
unemp_yoy_change	<code>unemployment_rate - unemp_lag4</code>	Year-over-year labor trend
unemp_momentum_4q	Mean slope of last 4 values (trend direction)	Positive = worsening market; negative = improving
unemp_volatility_8q	Rolling 8-quarter std	Long-term labor market instability
unemp_acceleration	<code>unemp_qoq_change - (unemp_lag1 - unemp_lag2)</code>	Measures if unemployment is worsening faster



## B. Relative & Cross-Market Context

Real estate investors always benchmark cities against peers.

Feature Name	Formula / Logic	Meaning
<b>unemp_above_national</b>	Local – National unemployment	Relative economic strength
<b>unemp_above_region</b>	Local – Regional average	Captures regional competitiveness
<b>market_unemp_zscore</b>	(Local – Mean) / Std (within market)	Normalized deviation from historical norm
<b>market_unemp_rank</b>	Rank of market's unemployment in that quarter	Market competitiveness indicator
<b>unemp_spatial_neighbor_avg</b>	Average unemployment of neighboring markets	Spatial spillover / contagion risk
<b>unemp_neighbor_gap</b>	Local – Neighbor average	How insulated or exposed a market is

## C. Supply-Demand Pressure Features

Top-tier firms always combine labor, supply, and vacancy fundamentals.

Feature Name	Formula / Logic	Meaning
<b>supply_pressure</b>	New units completed / Existing stock	Local construction intensity
<b>unemp_x_supply_pressure</b>	<b>unemployment_rate</b> * <b>supply_pressure</b>	Captures weak demand + high supply risk
<b>absorption_rate</b>	Leased space / New completions	Market health proxy
<b>unemp_x_absorption</b>	Interaction	Vacancies rise faster when both weak demand and high unemployment occur
<b>inventory_growth_4q</b>	$\Delta$ total stock over last 4 quarters	Supply growth trend

## D. Vacancy-Labor Interactions

Used to model feedback loops (vacancy ↔ labor).

Feature Name	Formula / Logic	Meaning
<b>vacancy_lag1, vacancy_lag2</b>	Prior vacancy	Persistence of vacancy cycles
<b>unemp_x_vacancy_lag1</b>	Interaction	Whether high unemployment worsens existing high vacancy
<b>vacancy_momentum_4q</b>	4-quarter slope	Direction of vacancy trend
<b>vacancy_shock_flag</b>	1 if $\Delta$ vacancy > 95th percentile	Stress indicator
<b>vacancy_gap_to_trend</b>	Current vacancy – rolling 4-quarter mean	Above/below structural level

## E. Macro & Policy Regime Signals

Sophisticated real estate models use these as “macro controls.”

Feature Name	Formula / Logic	Meaning
interest_rate_level	e.g., Fed funds or 10-yr treasury	Financing cost proxy
credit_availability_index	External data	Lending conditions
recession_flag	1 if national_unemp > 7%	Macro regime indicator
policy_change_dummy	1 if quarter coincides with major tax or zoning reform	Structural break marker
inflation_interaction	unemployment_rate * inflation_rate	Stagflation or overheating context

## F. Economic & Demographic Synergies

Adds realism to your unemployment–vacancy link.

Feature Name	Formula / Logic	Meaning
population_growth_rate	YoY change in population	Demand driver
migration_inflow_rate	New residents / total population	Vacancy absorber
unemp_x_population_growth	Interaction	Combines labor + demand shifts
median_income_growth	Quarterly change in income	Purchasing power shift
affordability_index	Rent / Income	Stress indicator

## G. Derived Indices & Composite Indicators

These are used by Moody's and CBRE in market outlook models.

Feature Name	Formula / Logic	Meaning
labor_market_stress_index	Weighted z-score of unemployment, volatility, and slope	Summary of job market tightness
demand_supply_ratio	Absorption / New supply	Key commercial metric
economic_health_score	PCA of unemployment, GDP, income growth	Latent macro strength factor
vacancy_pressure_index	(Unemp + supply_pressure) – (absorption + pop_growth)	Combined imbalance signal

## H. Implementation Skeleton (Python / Pandas)

```
python

import pandas as pd
import numpy as np

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

# --- Temporal features
df['unemp_lag1'] = g['unemployment_rate'].shift(1)
df['unemp_lag2'] = g['unemployment_rate'].shift(2)
df['unemp_lag4'] = g['unemployment_rate'].shift(4)
df['unemp_qoq_change'] = df['unemployment_rate'] - df['unemp_lag1']
```

```

df['unemp_yoy_change'] = df['unemployment_rate'] - df['unemp_lag4']
df['unemp_volatility_8q'] = g['unemployment_rate'].rolling(8).std().reset_index(level=0,
drop=True)

# --- Relative to national
national_unemp = df.groupby('quarter')['unemployment_rate'].mean()
df['unemp_above_national'] = df['unemployment_rate'] - df['quarter'].map(national_unemp)

# --- Vacancy dynamics
df['vacancy_lag1'] = g['vacancy_rate'].shift(1)
df['unemp_x_vacancy_lag1'] = df['unemployment_rate'] * df['vacancy_lag1']



# --- Example interaction with supply
df['unemp_x_supply_pressure'] = df['unemployment_rate'] * df['supply_pressure']

# --- Composite example
df['vacancy_pressure_index'] = (
    df['unemployment_rate'] + df['supply_pressure']
    - (df['absorption_rate'] + df['population_growth_rate'])
)

```

## Next Step Options

Would you like me to:

1.  Generate a **complete Python feature-engineering function or pipeline** that automatically builds all these features for your panel data; or
2.  Provide a **visual diagnostic notebook** showing which of these engineered unemployment features have the strongest relationship with vacancy\_rate across markets?

Which would be most useful for your current stage?



So me some more features engineering with unemployment\_rate.

Perfect 🍌 — now we're entering **"elite-tier" macro-real-estate feature engineering**, where you extract as much signal as possible from `unemployment_rate` by combining **economic logic**, **time-series behavior**, and **cross-market intelligence**.

Let's go beyond the typical lags and differences — we'll focus on features that top quant teams (Moody's Analytics, CBRE Econometric Advisors, Blackstone RE, etc.) often build to explain **vacancy**, **rents**, and **pricing cycles** using unemployment data.

## 1. Structural & Regime-Sensitive Features

Feature Name	Definition	Intuition
<b>unemp_trend_strength</b>	Slope of unemployment over last 8 quarters (via linear regression)	Captures structural improvement or deterioration — slow and persistent shifts matter most to vacancy.
<b>unemp_phase</b>	Encoded as "rising", "falling", or "stable" regime (using thresholds on slope)	Vacancy reacts differently in recovery vs contraction phases.

Feature Name	Definition	Intuition
<b>unemp_recession_flag</b>	1 if unemployment > (long-term avg + $1.5\sigma$ )	Marks local labor recession phases.
<b>unemp_recovery_speed</b>	$(\text{unemp\_lag4} - \text{unemployment\_rate}) / 4$	Average quarterly decline in unemployment → how fast recovery is happening.
<b>unemp_cycle_position</b>	Phase angle from cyclical decomposition (e.g., from Hodrick-Prescott filter)	Places market in a business cycle position: boom ↔ bust.
<b>unemp_shock_magnitude</b>		$\Delta$ unemployment
<b>unemp_stability_score</b>	$1 / \text{rolling\_std}(8q)$	Inverse volatility; stable job markets support lower vacancy.

## 2. Derivative & Momentum Features (Fine-Grained Time-Series Signals)

Feature	Formula	Meaning
<b>unemp_first_derivative</b>	$\Delta \text{unemployment} = U_t - U_{(t-1)}$	Short-term direction of labor market.
<b>unemp_second_derivative</b>	$\Delta^2 \text{unemployment} = \Delta_t - \Delta_{(t-1)}$	Acceleration or deceleration of job loss/recovery.
<b>unemp_sma_ratio</b>	$U_t / \text{rolling\_mean}(4q)$	Measures deviation from short-term trend.
<b>unemp_ema_diff</b>	$U_t - \text{EMA}(U, \text{span}=4)$	Captures how far current unemployment is from its exponential average.
<b>unemp_trend_reversal_flag</b>	1 if $\text{sign}(\Delta_t) \neq \text{sign}(\Delta_{(t-1)})$	Detects turning points — valuable for predicting vacancy reversals.
<b>unemp_skewness_8q</b>	Rolling skew of last 8 quarters	Detects asymmetric cycles (steep up / slow down).

## 3. Cross-Market / Spatial Intelligence

Feature	Description	Why It Helps
<b>unemp_diff_from_neighbors</b>	Local $U - \text{Mean}(U \text{ of neighboring markets})$	Detects geographic labor divergences that precede vacancy shifts.
<b>unemp_neighbor_volatility</b>	Std of $U$ in nearby markets	Labor instability nearby often spills over regionally.
<b>unemp_cluster_mean</b>	Mean unemployment for cluster of similar markets (industry, population)	Captures peer-group effect on vacancy.
<b>unemp_spillover_index</b>	Weighted avg of neighbor unemployment (weighted by distance or trade flows)	Measures cross-market exposure.

(You can build neighbor networks using K-NN on geography or economic similarity.)

## 4. Composite & Hybrid Indices (Used by Institutional Analysts)

Index	Formula	Interpretation
<b>labor_stress_index</b>	$(\text{zscore}(U\_level) + \text{zscore}(U\_volatility) + \text{zscore}(U\_trend\_strength)) / 3$	Summary measure of labor pressure — higher → more stress.
<b>market_health_index</b>	$1 - \text{normalized}(\text{labor\_stress\_index})$	Inverse of above; higher = healthier market.
<b>shock_adjusted_unemp</b>	$U\_t + 0.5 \times \text{unemp\_shock\_magnitude}$	Adjusts unemployment for recent shock impact.
<b>employment_gap</b>	Potential_employment – Actual_employment (if employment data available)	Long-run slack measure; persistent gap means structural vacancy risk.

## 5. Interaction Features (Real Power Features)

Combine unemployment with other variables you already have in your dataset.

Feature	Formula / Example	Why It Matters
<b>unemp_x_rent_growth</b>	$U \times \text{rent\_growth}$	Links labor weakness to rent pressure.
<b>unemp_x_new_supply</b>	$U \times \text{new\_supply\_rate}$	Reveals vulnerability to oversupply when jobs are scarce.
<b>unemp_x_inflation</b>	$U \times \text{inflation\_rate}$	Models stagflation environments.
<b>unemp_change_x_construction_cost</b>	$\Delta U \times \text{construction\_cost}$	High unemployment + rising costs = delayed projects, affecting vacancy lag.
<b>unemp_x_interest_rate</b>	$U \times \text{interest\_rate}$	Captures capital market tension effects.
<b>unemp_x_pop_growth</b>	$U \times \text{population\_growth\_rate}$	If jobs fall while population grows, vacancy may spike sharply.

## 6. Transformational Features (Nonlinear Transformations)

Feature	Definition	Insight
<b>log_unemp</b>	$\log(1 + U)$	Compresses scale, reduces skewness.
<b>inv_unemp</b>	$1 / (1 + U)$	Highlights improvements during recovery.
<b>unemp_squared</b>	$U^2$	Captures nonlinear vacancy sensitivity (e.g., large jumps hurt disproportionately).
<b>sqrt_unemp</b>	$\sqrt{U}$	Smooths small fluctuations while keeping shape.
<b>zscore_unemp_global</b>	$(U - \text{global\_mean}) / \text{global\_std}$	Comparable across markets.

## 7. Forward-Looking (Predictive) Features

Feature	How to Compute	Rationale
<b>predicted_unemp_next_q</b>	Use ARIMA/LightGBM to forecast next quarter unemployment	Vacancy often depends on <i>expected</i> unemployment.



Feature	How to Compute	Rationale
forecast_error_unemp	(Actual – Predicted)	Surprise component → stronger short-term effect.
expected_trend_2q	Projected change over 2 quarters	Forward economic signal.
unemp_diff_leading_vacancy	Correlation of $U_{(t-1:t-2)}$ with $vacancy_t$	Used as a learned feature for dynamic lags.

## 8. Sector-Specific Variations (Real Estate Focus)

Sector	Example Feature	Description
Office	unemp_white_collar_rate	Better proxy than total unemployment.
Retail	unemp_services_sector	Retail vacancy follows consumer-facing job losses.
Industrial	unemp_logistics_index	Derived from transport/manufacturing employment.
Apartments	unemp_young_population	Youth job market drives rental demand.

(You can get sectoral unemployment data from BLS or similar sources.)

## 9. Example Python Snippet for a Few of These

```
python

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

# Rolling trend via linear regression (8-quarter)
def rolling_trend(series, window=8):
    trend = series.rolling(window).apply(
        lambda x: np.polyfit(range(len(x)), x, 1)[0], raw=False
    )
    return trend

df['unemp_trend_strength'] = g['unemployment_rate'].apply(rolling_trend)
df['unemp_recovery_speed'] = (g['unemployment_rate'].shift(4) - df['unemployment_rate']) / 4
df['unemp_stability_score'] = 1 / g['unemployment_rate'].rolling(8).std()
df['unemp_acceleration'] = (
    df['unemployment_rate'] - g['unemployment_rate'].shift(1)
) - (g['unemployment_rate'].shift(1) - g['unemployment_rate'].shift(2))
df['unemp_phase'] = np.select(
    [
        df['unemp_trend_strength'] > 0.1,
        df['unemp_trend_strength'] < -0.1
    ],
    ['rising', 'falling'],
    default='stable'
)
```

## 10. Combine These Intelligently

When building your model:

- Use **time-aware cross-validation** (since these are sequential features).
- Standardize by **market** for interpretability.

- Track feature importance using **SHAP** — you'll see unemployment's lag, volatility, or interactions emerge as top drivers.

---

If you'd like, I can now:

1. 🔧 Generate a **feature-engineering module (Python class)** that constructs *all these unemployment-based features automatically*,
2. 🧩 Or, show you **which combinations historically yield strongest lift** in Moody's-style vacancy models (based on real research patterns).

Which direction would you prefer next — the **code module** or the **research-backed shortlist of most predictive features**?