

Utilities employment 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 `utilities_employment` and response `vacancy`. Help me understand how it works

Perfect — let's break this down carefully and connect **"utilities_employment"** to **"vacancy"** within your **real estate dataset** context.



1. What is *Utilities Employment*?

Utilities employment measures the number of people employed in the **utilities sector**, which includes:

- Electricity generation and distribution
- Natural gas supply
- Water and wastewater management
- Waste management services




It is typically **reported per region (e.g., metro, city, or state)** and **changes quarterly** depending on infrastructure projects, population growth, and economic health.



2. Why It Matters for Real Estate

Utilities employment acts as a **proxy for infrastructure demand and regional economic activity**.

Let's see how it connects to *vacancy rates* (your response variable):

Scenario	What Happens in Utilities Employment	Impact on Vacancy
 Rising Utilities Employment	Indicates infrastructure expansion and growth in service capacity → signals population or business growth	↓ Vacancy tends to decrease (higher occupancy)
 Falling Utilities Employment	Indicates reduced infrastructure investment, industrial slowdown, or migration outflow	↑ Vacancy tends to increase (less demand)
 Stable Utilities Employment	Indicates steady economic conditions and consistent service coverage	Vacancy remains stable

So, it **indirectly reflects regional growth** — more utility workers → more projects → higher housing and office demand.



3. How to Quantify the Relationship (for modeling)

You can create both **level** and **change-based** features from `utilities_employment`.

◆ a. Level Features

1. `utilities_employment` (raw value)

- Shows absolute scale of employment.

2. `log_utilities_employment`

- Log transformation can stabilize variance across markets.

◆ b. Growth / Momentum Features

1. Quarter-over-Quarter (QoQ) Change

$$\Delta_{QoQ} = \frac{UE_t - UE_{t-1}}{UE_{t-1}}$$

Measures short-term change; useful for detecting expansions or contractions.

2. Year-over-Year (YoY) Change

$$\Delta_{YoY} = \frac{UE_t - UE_{t-4}}{UE_{t-4}}$$

Helps track structural trends rather than seasonal variation.

3. Moving Average (MA_4, MA_8)

Smooths cyclical fluctuations; reveals medium-term trends.

4. Deviation from Trend

Using something like HP filter or linear trend deviation:

$$UE_{gap} = UE_t - \text{trend}(UE_t)$$

Positive gap → expansion phase.



4. Economic Intuition by Property Type

Property Type	Why Utilities Employment Matters	Expected Effect
Residential / Apartments	Growth in utilities signals expansion of basic services → supports new housing projects	Vacancy ↓
Industrial	Utilities employment rises with manufacturing & energy demand	Vacancy ↓
Office	Indirect link — infrastructure improvement attracts more firms	Vacancy ↓ (moderately)
Retail	Utilities expansion indicates growing local consumption capacity	Vacancy ↓



5. Advanced Feature Engineering Ideas

Here's what **top analysts** in real estate use:

Interaction Features

- **utilities_employment × population_growth**
→ Captures synergy between infrastructure and demographics.
- **utilities_employment × manufacturing_employment**
→ Reveals industrial co-movement patterns.

Lead/Lag Features

- **lag_1, lag_2, ...** of utilities employment
(utilities sector often **leads** real estate activity by 1–2 quarters)
- Try correlation between **vacancy_t** and **utilities_employment_{t-1}**

Market Normalization

- **utilities_employment / total_employment**
→ Relative strength of the utilities sector in the regional economy.

6. Expected Empirical Relationship

In most U.S. markets, you will observe:

- **Negative correlation** between **utilities_employment_growth** and **vacancy**.
- Stronger effect in **industrial** and **apartment** markets.
- Weaker or lagged effects in **office** and **retail** segments.

Would you like me to help you **design out-of-the-box feature engineering ideas** (like how Moody's or REIS Analytics teams might construct second-order and cross-market features) specifically for **utilities_employment**?

That would include advanced interactions, lag structures, and spatial spillover features.



Help with out of box features engineering for this column.

Excellent — since you're working on a **real estate forecasting model** (vacancy as response) and want **out-of-box feature engineering** for **utilities_employment**, let's go beyond basic transformations and move toward **what top-tier real estate analytics teams (like Moody's, CoStar, REIS, CBRE Econometrics)** actually do.

We'll divide this into layers:

1. Temporal Dynamics Features (Time-pattern engineering)

Utilities employment reflects infrastructure cycles that often **lead** property market activity by 1–3 quarters. Capture those timing effects:

♦ Lag & Lead Features

Feature	Description	Economic logic
<code>utilities_employment_lag1, lag2, lag4</code>	Past values	Utility growth leads vacancy tightening
<code>utilities_employment_lead1, lead2</code>	Forward values	Can test reverse causality or infrastructure response
<code>Δutilities_employment_QoQ,</code> <code>Δutilities_employment_YoY</code>	Growth rates	Captures expansion/contraction momentum
<code>rolling_avg_4q, rolling_std_4q</code>	1-year trend and volatility	Steady vs unstable employment patterns

→ **Use case:** Markets where utility jobs grow steadily show sustained occupancy gains.

2. Cyclical Positioning Features

Utilities employment is cyclical — linked with industrial production and construction. Capture its **phase** and **cycle position**.

Feature	Formula / idea	Interpretation
<code>cycle_phase = (utilities_employment - rolling_mean_8q) / rolling_std_8q</code>	Z-score vs recent trend	Positive = expansion phase
<code>trend_strength = rolling_mean_4q / rolling_mean_8q</code>	Momentum ratio	>1 = growth accelerating
<code>turning_point_flag</code>	1 if sign change in growth	Detects cycle peaks/troughs

→ These help identify **when** a market is shifting — critical for early vacancy prediction.

3. Structural/Contextual Features

Utilities employment represents essential services. You can contextualize it relative to broader economic variables:

Feature	Description	Why it matters
<code>utilities_share = utilities_employment / total_employment</code>	Utilities sector's share	High share = infrastructure-heavy economy
<code>utilities_per_capita = utilities_employment / population</code>	Workforce density	Infrastructure support intensity
<code>utilities_vs_industrial_ratio = utilities_employment / industrial_employment</code>	Relative strength to industrial base	Industrial synergy vs saturation
<code>utilities_vs_construction_ratio = utilities_employment / construction_employment</code>	Infrastructure relative to building	Indicates service adequacy

→ These ratios reveal **how balanced or strained** a region's infrastructure employment is relative to its economy.

4. Cross-Market / Spatial Spillover Features

Utilities investment often benefits **neighboring markets** (e.g., power grid upgrades spanning metro areas).

You can model **spatial dependencies** using weighted averages:

Feature	Description
<code>neighbor_avg_utilities_employment</code>	Weighted average of nearby markets' utilities jobs
<code>neighbor_growth</code>	Average QoQ growth of nearby markets' utilities employment
<code>diff_from_neighbor = utilities_employment - neighbor_avg_utilities_employment</code>	Relative over/underinvestment

→ Capture **regional spillover**—markets with underdeveloped utilities may face future vacancy increases as firms relocate.

5. Interaction Features

Combine utilities employment with demand, pricing, or supply-side metrics to capture **nonlinear dependencies**:

Interaction	Economic intuition
<code>utilities_employment × population_growth</code>	Population growth amplifies infrastructure needs
<code>utilities_employment × manufacturing_employment</code>	Industrial co-dependence—power demand drives space demand
<code>utilities_employment × rent_growth</code>	Where utilities expand while rent grows → strong absorption
<code>utilities_employment × vacancy_lag</code>	Infrastructure growth moderates previous vacancy shocks

→ In practice, Moody's-type models include these cross-term effects to model *local resilience and lagged adjustment*.

6. Stability and Shock Features

Utilities employment is usually stable, so sudden deviations can indicate **economic or natural shocks**:

Feature	Description
<code>shock_flag = 1 if abs(ΔQoQ) > 2×rolling_std_4q</code>	Detects big infrastructure disruptions
<code>shock_magnitude = abs(ΔQoQ)</code>	Measures severity
<code>shock_persistence = count_consecutive_quarters_of_decline</code>	Tracks duration of downturns

→ Use for modeling **abnormal vacancy changes** (e.g., power plant closure → spike in industrial vacancy).

7. Long-term / Structural Alignment

Track how utilities growth aligns with long-term demand fundamentals:

Feature	Description
<code>10yr_trend_slope</code>	Linear regression slope over past 10 years
<code>trend_vs_population = trend_slope_utilities / trend_slope_population</code>	Infrastructure adequacy
<code>trend_vs_gdp = trend_slope_utilities / trend_slope_gdp</code>	Economic scaling balance

→ This gives insight into whether utility capacity is keeping up with regional expansion (a major signal for developers).

8. Market Clustering Feature (Macro-Signal)

You can cluster markets by utilities employment dynamics:

- Run KMeans or PCA on `utilities_employment_growth`, `volatility`, `trend_strength`
- Assign a **cluster ID feature** (e.g., "Infrastructure Growth Cluster 3")
- Use this as a categorical explanatory variable — different clusters behave differently in vacancy cycles.

9. Derived Signal Ratios (used by top analysts)

Feature	Formula	Interpretation
<code>utilities_employment_to_vacancy_ratio</code>	<code>utilities_employment / vacancy</code>	Inverse infrastructure strain
<code>utilities_employment_diff = utilities_employment - utilities_employment_lag4</code>	Yearly delta	
<code>utilities_employment_index = utilities_employment / base_year_value</code>	Indexed trend	

10. Machine-Learning-Friendly Transformations

To help nonlinear models capture subtle relationships:

- Polynomial terms (degree 2 or 3)
- Log-differences (`log(UE_t) - log(UE_{t-1})`)
- Winsorized values to reduce volatility
- Percentile rank by market (e.g., top 10% infrastructure-heavy metros)

Example:

Suppose Atlanta's utilities employment rose steadily over 6 quarters while Austin's declined slightly:

- Atlanta's **lagged utilities growth** will correlate strongly with **lower future vacancy**.
- Austin's **negative growth momentum** combined with **high vacancy lag** may predict persistent softness.

Would you like me to now create a **Python-style feature engineering blueprint** (with pseudo-code or pandas transformations) showing how to compute all these categories efficiently from your quarterly panel data?

That would make it easier for you to directly plug into your modeling pipeline.



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're asking for **truly top-tier, cutting-edge feature engineering ideas** for **utilities_employment** as used by the **leading real estate analytics firms** (Moody's, CoStar, REIS, CBRE Econometrics). Let's go beyond the standard lag/growth features and focus on **sophisticated, industry-grade signals** that capture macro, micro, and spatial dynamics for **vacancy forecasting**.

Advanced Temporal & Cyclical Features

Top analysts know that utilities employment **precedes property demand shifts**, so they model cycles precisely:

1. Phase-Aware Features

- Use **Hodrick-Prescott (HP) filter** or STL decomposition to separate trend vs cycle.
- Feature examples:
 - `utilities_cycle = utilities_employment - trend`
 - `cycle_amplitude = max_cycle - min_cycle`
- Captures **expansion or contraction phase** for vacancy predictions.

2. Adaptive Lag Features

- Compute **cross-correlation between utilities_employment and vacancy** to find optimal lag per market.
- Features:
 - `utilities_lag_optimal` → use lag where correlation with vacancy is strongest.
- This ensures model captures **market-specific lead/lag effects**.

3. Volatility & Stability Metrics

- Quarterly fluctuations often predict **rental or vacancy instability**.

- Features:
 - `rolling_std_4q` → short-term volatility
 - `rolling_std_8q / rolling_mean_8q` → coefficient of variation
 - `shock_indicator` → large quarter-over-quarter jumps

2 Spatial & Market Interactions

Utilities investments often affect neighboring markets—analysts use **spatial spillover features**:

1. Neighboring Market Averages

- `neighbor_avg_utilities_employment` → weighted by geographic proximity or economic linkages.
- Captures **regional infrastructure momentum**.

2. Spatial Relative Strength

- `utilities_relative_to_region = utilities_employment / neighbor_avg_utilities_employment`
- Markets under or over-invested relative to peers often show divergent vacancy trends.

3. Cross-Sector Interactions

- Combine utilities with manufacturing, construction, and tech employment.
- Example:
 - `utilities_employment × construction_employment` → predicts industrial/commercial vacancy
 - `utilities_employment × population_growth` → predicts residential vacancy

3 Normalized & Structural Features

Industry leaders focus on **ratios and structural signals** rather than absolute numbers:

1. Per Capita & Per Job

- `utilities_per_capita = utilities_employment / population`
- `utilities_share_of_total_employment = utilities_employment / total_employment`
- These normalize across markets with different sizes.

2. Trend Alignment Ratios

- Compare utilities growth vs GDP, manufacturing, or construction:
 - `utilities_vs_gdp_growth_ratio`
 - `utilities_vs_manufacturing_growth_ratio`
- Deviations signal **over- or under-supply of infrastructure**, impacting future vacancy.

4 Leading Indicators / Predictive Signals

Top-tier models often create **forward-looking features**:

1. Infrastructure Momentum Index

- Weighted sum of utilities employment growth + permits + private construction activity.
- Captures **composite demand pressure** before vacancy changes.

2. Acceleration Features

- $\Delta(\Delta \text{Utilities_employment_QoQ}) \rightarrow$ second derivative (rate of change of growth)
- Detects **rapid expansions or contractions** which strongly predict vacancy shifts.

3. Structural Break Flags

- Detect sudden jumps (power plant shutdown, new grid investments) using:
 - $\text{abs}(\Delta \text{Utilities_employment}) > \text{threshold}$
- Models often treat these as special events rather than regular trends.

5 Market Segmentation & Cluster Features

Moody's and similar firms often **group markets by infrastructure dynamics**:

1. Cluster ID Features

- Cluster markets based on:
 - Utilities growth pattern
 - Volatility
 - Trend vs GDP alignment
- Assign cluster ID as a categorical variable. Example:
 - Cluster 1: High-growth stable utilities
 - Cluster 2: Volatile but high absolute employment
 - Cluster 3: Low-growth, underinvested markets
- Vacancy models adjust predictions based on cluster.

2. Percentile & Rank Features

- $\text{utilities_employment_rank} = \text{percentile rank within region}$
- Detects whether a market is **top/bottom quartile** in infrastructure, which often correlates with absorption rates.

6 Interaction with Macro & Policy Signals

Top analysts combine utilities with macroeconomic or regulatory variables:

- $\text{utilities_employment} \times \text{interest_rate_change}$
- $\text{utilities_employment} \times \text{zoning_permit_index}$
- $\text{utilities_employment} \times \text{utility_capex} / \text{population}$
- These capture **policy or macro shocks** that impact property demand before it appears in vacancy.

7 Machine Learning-Friendly Transformations

To let tree-based or deep models detect subtle patterns:

- **Polynomial Features:** $(\text{utilities_employment})^2$, $(\text{utilities_employment_growth})^2$
- **Log-Differences:** $\log(\text{UE}_t) - \log(\text{UE}_{t-1})$ to reduce skew
- **Winsorization:** Cap extreme spikes to avoid noise
- **Interaction Terms:** automatically generated via pipelines (e.g., $\text{UE} \times \text{Manufacturing} \times \text{Construction}$)

✓ Summary of “Top-Notch” Philosophy

1. Don't rely only on raw employment numbers—use **normalized, lagged, and growth-based signals**.
2. Capture **phase and cyclical position** of infrastructure trends.
3. Include **spatial and cross-market interactions**.
4. Model **shock, volatility, and acceleration** to detect leading vacancy signals.
5. Cluster markets and use macro interactions for structural understanding.
6. Feed ML-friendly transformations to capture **nonlinear relationships**.

If you want, I can create a **ready-to-implement feature engineering template for `utilities_employment`** exactly in **Python/pandas style**, covering all these advanced industry-standard ideas. This is exactly how top real estate analytics teams structure their input data.

Do you want me to do that next?



So me some more features engineering with `utilities_employment`.

Perfect — let's go even deeper into **advanced, out-of-the-box features** for `utilities_employment`, the kind **top real estate analytics teams like Moody's or CoStar** would create for vacancy forecasting. I'll give you **categories of features** beyond the usual lags, growth rates, and rolling averages.

1 Dynamic / Temporal Features

These capture how `utilities_employment` changes over time, including momentum and acceleration:

- **Quarterly Momentum:** $\text{UE}_t - \text{UE}_{t-1}$
- **Yearly Momentum:** $\text{UE}_t - \text{UE}_{t-4}$
- **Acceleration:** $(\text{UE}_t - \text{UE}_{t-1}) - (\text{UE}_{t-1} - \text{UE}_{t-2}) \rightarrow$ detects rapid hiring/firing cycles
- **Rolling Trend Slope:** linear regression slope of the past 4–8 quarters \rightarrow captures long-term trend

- **Exponential Weighted Moving Average (EWMA):** gives more weight to recent quarters to capture short-term demand signals

2 Relative / Normalized Features

Normalize utilities employment to make it comparable across markets:

- **utilities_share** = $UE / total_employment$ → how important utilities sector is in this market
- **utilities_per_capita** = $UE / population$ → workforce intensity per person
- **UE_growth_relative** = $(UE_t - UE_{t-1}) / total_employment_growth$ → growth relative to market size

3 Interaction Features

Combine utilities employment with other economic or real estate indicators:

- **UE × construction_employment** → synergy of infrastructure and building activity
- **UE × population_growth** → predicts residential vacancy trends
- **UE × industrial_employment** → predicts industrial vacancy absorption
- **UE × lagged_vacancy** → captures moderating effect of infrastructure on past vacancy

4 Cyclical / Phase Features

Utilities employment follows **economic cycles**, which affects vacancy:

- **Cycle deviation:** $UE - rolling_mean_8q$ → positive = expansion, negative = contraction
- **Cycle volatility:** rolling standard deviation over 4–8 quarters
- **Phase indicators:** 1 if $\Delta UE > rolling_mean_growth$ else 0 → indicates entering growth phase

5 Spatial / Market Comparison Features

Real estate markets don't exist in isolation:

- **UE_relative_to_neighbors** = $UE / avg(UE_neighboring_markets)$ → over- or under-invested regions
- **neighbor_growth** = $avg(\Delta UE_neighboring_markets)$ → spillover effect on vacancy
- **market_cluster_ID** → cluster markets based on UE dynamics and trend patterns

6 Shock & Volatility Features

Utilities employment shocks often predict sudden changes in vacancy:

- `shock_flag = 1` if `abs(Δ UE) > 2 * rolling_std_4q`
- `shock_duration` = consecutive quarters of declining UE
- `shock_magnitude` = `abs(Δ UE)`

7 Predictive / Leading Indicators

Utilities employment often **leads vacancy changes**:

- `lagged_UE_1q`, `lagged_UE_2q`, `lagged_UE_3q` → best lag determined by cross-correlation with vacancy
- `UE_growth_lead` = `Δ UE_t+1` → anticipates upcoming vacancy shift

8 Trend vs Macro Alignment

Compare UE trend to broader economic signals:

- `UE_vs_population_trend` = `UE_slope / population_slope` → shows infrastructure keeping pace with population
- `UE_vs_GDP_trend` = `UE_slope / GDP_slope` → shows regional infrastructure adequacy

9 Machine Learning Transformations

For nonlinear models:

- Polynomial terms: `(UE)^2`, `(Δ UE)^2`
- Log transformations: `log(UE + 1)` → stabilizes variance
- Percentile ranks: `UE_rank_in_market` → identifies top vs bottom quartile markets

✓ Summary

Top-notch features from `utilities_employment` are about **capturing timing, scale, phase, volatility, interactions, spatial context, shocks, and macro alignment**. These features make ML models more robust at predicting vacancy because they encode **economic reality**, not just raw counts.

If you want, I can **create a concrete Python/pandas pipeline** for `utilities_employment` that **computes all these features automatically** for your quarterly panel dataset. It would be exactly like Moody's-style feature engineering.

Do you want me to do that?