

Feature response analysis



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

Got it! Let's break this down carefully. You are working with **Moody's real estate dataset**, which is **quarterly, multi-market, multi-feature**, and you want to understand the relationship between:

- **Feature:** `transportation_warehousing_employment`
- **Response:** `vacancy` (probably the vacancy rate in commercial or industrial properties)

I'll explain step by step how this works and what you should consider.

1. Understanding the Feature: `transportation_warehousing_employment`

- This represents **employment in the transportation and warehousing sector** within a particular market and quarter.
- It's usually measured as **number of jobs or employment index**.
- Sectors included:
 - Logistics, freight, and shipping
 - Warehousing and storage
 - Trucking and delivery services
 - Transportation support activities

Why it matters for real estate:

- High employment in transportation & warehousing can indicate **demand for industrial space** like warehouses, distribution centers, and logistic hubs.
- Growth in this sector usually **reduces vacancy rates** in industrial or logistic properties.
- Conversely, a decline might lead to **higher vacancy**, signaling less demand.

2. Understanding the Response: `vacancy`

- Vacancy is typically the **percentage of unoccupied space** in a particular property type (office, retail, industrial, etc.) in a market.
- It's a key indicator of **market health and demand-supply balance**.
- A higher vacancy → less demand or oversupply

- A lower vacancy → strong demand, tight market

3. Expected Relationship

Since this feature is tied to the **employment that directly uses industrial/transport spaces**, we can make some educated hypotheses:

Scenario	transportation_warehousing_employment	Vacancy Rate	Interpretation
Employment ↑	Likely ↑	Vacancy ↓	More workers → more activity → more space leased → lower vacancy
Employment ↓	Likely ↓	Vacancy ↑	Less activity → less demand for industrial space → higher vacancy

So, in general, you expect a **negative correlation** between **transportation_warehousing_employment** and industrial vacancy rates.

4. Things to Consider in Analysis

Because your data is **quarterly and multi-market**, simple correlation may not tell the full story:

1. Lag effect:

- Employment change may not immediately impact vacancy.
- For example, it may take **1-2 quarters** for a new warehouse to be leased after hiring picks up.

2. Market differences:

- NYC vs. Austin vs. Atlanta might have very different demand dynamics.

3. Normalization / per capita adjustments:

- Comparing raw employment numbers across markets may be misleading due to market size differences.
- Consider **employment density per 1000 people** or **market size adjustment**.

4. Interaction with other features:

- Industrial vacancy might also depend on:
 - Total industrial completions (**industrial_completions**)
 - Economic growth indicators (**GDP, manufacturing_employment**)
 - Other employment sectors (**professional_services_employment**)

5. Trend and seasonal effects:

- Transportation and warehousing may have **seasonal spikes** (holiday shipping, Q4).
- Quarterly data → check for **seasonality adjustment**.

5. Feature Engineering Ideas (Specific to This Feature)

1. Lagged Employment Features:

- `transportation_warehousing_employment_lag1` → employment from last quarter
- `transportation_warehousing_employment_lag2` → employment from two quarters ago

2. Growth / Change Rate:

- Quarter-over-quarter % change:

$$growth = \frac{employment_q - employment_{q-1}}{employment_{q-1}}$$

3. Market-Adjusted Employment:

- Normalize by total market employment or population to get relative size.

4. Rolling Average / Trend:

- 4-quarter rolling average to smooth out seasonal fluctuations.

5. Employment-to-Space Ratio (if space data exists):

- Compare employment growth to new warehouse completions in the same market.

6. Modeling Vacancy Using This Feature

- **Simple correlation:** to see if negative correlation exists.
- **Regression:**
 - `vacancy ~ transportation_warehousing_employment + other_features`
- **Time series / panel regression:**
 - Use lagged variables and market fixed effects to account for market-specific patterns.
- **Causal inference:**
 - If your goal is to **quantify impact**, you could use models like **difference-in-differences** if a big transportation hub opened in a city.

Key takeaway:

`transportation_warehousing_employment` is a leading indicator for industrial property vacancy. Higher employment generally signals **lower vacancy** in industrial/logistics spaces, but the effect may lag and vary by market. Properly engineered lag, growth, and normalization features can improve predictive power.

If you want, I can **draw a full schematic showing the relationship between `transportation_warehousing_employment`, lagged effects, and vacancy** for all property types—it makes this much easier to visualize.

Do you want me to do that?



Help with out of box features engineering for this column.

Perfect! Since you want **out-of-the-box, top-tier feature engineering ideas** for **transportation_warehousing_employment** (beyond standard lags and growth), here's a curated set that real estate/data science teams in top firms often use for industrial/logistics markets:

1. Advanced Lag & Trend Features

- **Lagged Employment Growth:**

Instead of simple lag, use **percentage growth over multiple quarters**:

$$\bullet \text{ tw_emp_qoq_growth} = (\text{TW_emp_q} - \text{TW_emp_q-1})/\text{TW_emp_q-1}$$

$$\bullet \text{ tw_emp_yoy_growth} = (\text{TW_emp_q} - \text{TW_emp_q-4})/\text{TW_emp_q-4}$$

This captures **short-term vs long-term trend effects** on vacancy.

- **Weighted Rolling Averages:**

Smooth volatility with **exponentially weighted moving averages (EWMA)**:

$$\bullet \text{ tw_emp_ewma} = 0.7*\text{TW_emp_q} + 0.3*\text{TW_emp_q-1}$$

Helps capture sustained employment trend rather than noisy spikes.

2. Market-Adjusted Features

- **Employment Share of Market:**

$$tw_{emp_share} = \frac{\text{TW}_{emp}}{\text{Total}_{market_{employment}}}$$

Standardizes across markets of different sizes.

- **Employment Density:**

$$tw_{emp_density} = \frac{\text{TW}_{emp}}{\text{industrial}_{space_sqft}}$$

Captures **jobs per square foot**, which often directly correlates with vacancy pressure.

- **Deviation from Market Trend:**

$$tw_{emp_dev} = \text{TW}_{emp} - \text{TW}_{mp_market_avg}$$

Highlights markets **over-performing or under-performing** relative to national/regional trend.

3. Interaction Features

- **Employment × Completions:**

$$\text{tw_emp_x_completions} = \text{TW_emp} * \text{industrial_completions}$$

Captures **supply-demand mismatch**: lots of new space + declining employment → vacancy rises.

- **Employment × Transportation Infrastructure Index:**

- If you have data on roads, ports, or airports, interaction captures the effect of accessibility on warehouse demand.

- **Employment × Other Sectors:**

- Combine with `manufacturing_employment` or `retail_employment` to reflect **downstream demand**.
-

4. Seasonal & Cyclical Features

- Seasonal Index:**

$tw_emp_seasonality = TW_{emp} / \text{avg}(TW_{emp_q_same_quarter_past_3_years})$

Removes regular seasonal peaks (e.g., Q4 shipping surge).

- Cyclic Momentum:**

Compare **current quarter growth vs previous 2-4 quarters** to detect acceleration/deceleration in industrial employment:

$$tw_{emp_momentum} = TW_{emp_growth_q} - TW_{emp_growth_{q-1}}$$

5. Derived Ratios

- Employment-to-Vacancy Ratio:**

$$tw_{emp_vacancy_ratio} = TW_{emp}/vacancy$$

Gives **stress on vacancy** relative to workforce.

- Change-in-Ratio:**

$$tw_{emp_vacancy_ratio_change} = (TW_{emp}/vacancy)_q - (TW_{emp}/vacancy)_{q-1}$$

Measures **pressure on market over time**.

- Relative to Industrial Space Completions:**

$$tw_{emp_space_ratio} = TW_{emp}/industrial_space_completed$$

High ratio → strong demand, low vacancy expected.

6. Anomaly / Event Features

- Significant Shifts Detection:**

Flag quarters with unusually high jumps or drops:

$$tw_{emp_anomaly} = 1 \text{ if } |TW_{emp_q} - TW_{emp_{q-4}}| > threshold$$

Could capture **new distribution center openings or closures**.

- Acceleration/Deceleration Flags:**

Binary flag when growth rate changes sign (from negative → positive, or vice versa), often preceding **vacancy trend changes**.

7. Market Cross-Effects (if multi-market modeling)

- Regional Spillover:**

Average employment growth of neighboring markets → accounts for **logistics hubs serving**

multiple regions.

- **Competition Index:**

Compare `TW_emp` to nearby markets' employment + available warehouse space to capture **demand shift risk**.

💡 Pro Tips

- Always combine **lagged, normalized, and interaction features**. These often outperform raw features in real estate.
- For quarterly data, consider **1-2 quarter lagged interactions** with completions or vacancy.
- Test **rolling windows of 2-8 quarters** for EWMA and momentum features.

If you want, I can **make a ready-to-use Python snippet** that engineers **10+ of these top-tier features automatically** for your `transportation_warehousing_employment` column. This will save a ton of time.

Do you want me to do that?



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

Ah, got it! You're looking for **next-level, "industry-grade" feature engineering** ideas that top real estate analytics teams actually use for employment-related features like `transportation_warehousing_employment` to predict vacancy. Let's go beyond the basics and go into **strategic, data-science-backed features**.

1. Leading Indicator Features

Top real estate teams know that employment in transport & warehousing often acts as a **leading indicator** of industrial demand.

- **Lagged Growth + Vacancy Forecast:**

- Create features like:

$$tw_{emp}lag1_growth = \frac{TW_{emp,q-1} - TW_{emp,q-2}}{TW_{emp,q-2}}$$

- Combine with vacancy:

$$tw_{emp}lag1_growth \times vacancy_{q-1}$$

- Captures **how prior employment growth is influencing current vacancy**.

- **Acceleration Indicators:**

- Measures second derivative:

$$tw_{emp_acceleration} = (TW_{emp,q} - TW_{emp,q-1}) - (TW_{emp,q-1} - TW_{emp,q-2})$$

- Detects **market inflection points**, used by investment analysts to anticipate tight supply.

2. Market Saturation / Absorption Features

- **Industrial Space Absorption Rate:**

$$absorption_rate = \frac{occupied_industrial_space}{total_industrial_space}$$

- Then create:

$$tw_{emp_to_absorption} = TW_{emp}/absorption_rate$$

- Shows **employment pressure relative to available space**.

- **Employment per Unit Industrial Space:**

$$emp_density = TW_{emp}/industrial_sqft$$

- Used by top analysts to **benchmark utilization efficiency**.

3. Supply-Demand Mismatch Features

- **Employment vs. Completions Ratio:**

$$tw_{emp_supply_ratio} = TW_{emp}/industrial_completions$$

- High ratio → low vacancy expected (high demand, low supply)
- Low ratio → potential oversupply → vacancy pressure.

- **Quarterly Change in Employment-to-Space Ratio:**

- Detects **short-term market shocks**, e.g., a new fulfillment center opening.

4. Seasonality & Cyclical Adjustment

- **Seasonally Adjusted Employment Index:**

- Divide by historical quarterly average for that market:

$$tw_{emp_sa} = TW_{emp,q}/avg(TW_{emp,q-4,q-8,q-12})$$

- Removes **recurring shipping/holiday cycles**.

- **Momentum vs. Historical Cycle:**

- Compare current growth to **long-term cycle trend**:

$$tw_{emp_momentum} = TW_{emp,q} - rolling_mean(TW_{emp,q-8:q-1})$$

- Top analysts use this to **anticipate market peak or trough**.

5. Interaction Features with Other Economic Drivers

- **Transportation Employment × GDP Growth:**
 - Captures how **industrial activity translates into property demand** across cycles.
- **Employment × Manufacturing Index:**
 - Many logistics hubs are tied to manufacturing.
 - Interaction reflects **derived demand for warehouse space**.
- **Employment × Retail/Wholesale Activity:**
 - Helps capture **e-commerce-driven demand spikes**.

6. Market Relative & Cross-Market Features

- **Relative Employment Growth vs National Average:**

$$rel_growth = TW_{emp,market}/TW_{emp,national_avg}$$

- Shows which markets are **outperforming peers**, influencing vacancy risk premium.

- **Neighboring Market Spillover:**

- For multi-market models, compute **weighted average of neighboring markets**:

$$tw_{emp_neighbor} = \sum_{i \in neighbors} TW_{emp,i} \times w_i$$

- Reflects **regional demand shifts**, often used by industrial RE analysts.

7. Shock/Event Features

- **Employment Shock Flags:**
 - Binary feature if quarter-over-quarter change exceeds threshold, e.g., $\pm 10\%$
 - Used to identify **new logistic hubs, port expansions, or closures**.
- **Cumulative Employment Shock:**
 - Sum of positive/negative shocks over last 4 quarters → detects **emerging trend**.

8. High-Level Real Estate Industry Tricks

1. Lagged Interaction with Vacancy:

$$tw_{emp_lag1} \times vacancy_lag1$$

- Captures **elasticity** of vacancy with respect to prior employment.

2. Rolling Rank / Percentile by Market:

- Rank this quarter's employment vs last 8–12 quarters → detect **overperformance or underperformance**.

3. Velocity of Change Features:

- Combine acceleration, deceleration, and momentum to capture **trend strength**.

4. Cluster-Based Features:

- Segment markets into clusters (high/medium/low industrial intensity) and compute cluster-relative employment index.

Key takeaway:

Top real estate analysts rarely rely on just raw employment numbers—they focus on **relative, lagged, interactive, momentum-based, and market-adjusted features** that capture both **trend and anomaly signals**. These are often the features that **give predictive edge** for vacancy forecasting.

If you want, I can make a **table of 15+ top-notch “ready-to-use” engineered features** for **transportation_warehousing_employment** with formulas and descriptions, exactly how Moody's or top RE analytics teams would implement them.

Do you want me to do that?



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

Perfect! Let's go **fresh**—completely new, realistic, and advanced feature ideas for **transportation_warehousing_employment** that top real estate/industrial analytics teams could use for vacancy modeling. I'll avoid repeating the earlier lags, growth, or rolling averages.

1. Employment Volatility Features

- **Quarterly Volatility (Std Dev)**

$$tw_{emp} volatility = std(TW_{emp,q-4:q})$$

- Measures how stable employment is. Highly volatile employment can **signal unstable industrial demand**, affecting vacancy risk.

- **Employment Spike Flag**

- Binary feature if quarterly change exceeds 1.5× standard deviation of last 4 quarters.
- Captures sudden expansions or contractions in logistics activity.

2. Employment Concentration / Distribution Features

- **Share of Employment in Total Industry Cluster**

$$tw_{emp_cluster_share} = TW_{emp}/\sum(TW_{emp,related_industries})$$

- Measures **how dominant transport & warehousing employment is** relative to other industrial sectors in that market.
- Helps predict vacancy in **specialized industrial hubs**.
- **Employment Herfindahl Index (HI)**

$$tw_{emp_HI} = \sum_i (TW_{emp,i}/total_industry_emp)^2$$

- Quantifies concentration of logistics employment vs other sectors.
- Highly concentrated markets may have **more sensitive vacancy to shocks**.

3. Employment Trend Shape Features

- **Slope of Employment Trend (Linear Fit)**

- Fit a line to last 4–8 quarters of employment and take slope:

$$tw_{emp_slope} = slope(quarters, TW_{emp})$$

- Captures whether employment growth is accelerating or decelerating **beyond simple difference**.

- **Convexity / Curvature of Trend**

- Fit a quadratic curve to last 4–8 quarters and extract curvature:

- Positive → accelerating growth
- Negative → slowing growth

- Indicates **upcoming tightening or loosening of vacancy**.

4. Employment-Driven Market Pressure Features

- **Employment Pressure Index (EPI)**

$$EPI = TW_{emp}/(available_warehouse_space + 1)$$

- Measures **pressure on existing industrial space**, high values → vacancy likely to drop.

- **Net Employment Absorption Index**

$$net_absorption = TW_{emp_growth} - industrial_space_growth$$

- Captures **relative mismatch between workforce growth and new supply**, a strong industrial real estate signal.

5. Employment Cycle Features

- **Cycle Phase Indicator**

- Compute ratio of current employment to peak/trough in last 8 quarters:
 - 0.9 → peak phase
 - <0.6 → trough phase
- Top analysts use this to **forecast industrial market tightness**.

- **Quarterly Employment Momentum Score**

$$momentum = (TW_{emp,q} - TW_{emp,q-1}) / (TW_{emp,q-1} - TW_{emp,q-2} + 1)$$

- Captures **acceleration relative to previous change**, often predictive of vacancy spikes.

6. Cross-Property Impact Features

- **Employment to Office/Industrial Mix Ratio**

$$tw_{emp_office_ratio} = TW_{emp_office} / office_employment$$

- Helps model how logistics employment interacts with other commercial property demand.

- **Warehouse Demand Score**

- Weighted sum of employment in transportation & warehousing + manufacturing + retail logistic support.
- Converts multi-sector employment into a **single industrial demand proxy**.

7. Regional / Spatial Features

- **Neighbor Market Employment Weighted Average**

$$tw_{emp_regional} = \sum_{i \setminus \text{neighbors}} TW_{emp,i} * weight_i$$

- Captures **spillover demand**; warehouses often serve regional markets, not just the local city.

- **Market Employment Rank**

- Rank the market vs all markets in transportation & warehousing employment.
- Useful for **relative market tightness indicators**.

8. Shock-Adjusted Features

- **Employment Shock Intensity**

$$shock_intensity = |TW_{emp,q} - median(TW_{emp,q-4:q-1})|$$

- Quantifies unusual jumps or drops; predictive of vacancy volatility.

- **Cumulative Positive/Negative Employment Shocks**

- Sum of all quarters where employment change > threshold.
 - Detects **emerging upward or downward trend** affecting warehouse demand.
-

 **Summary:**

These new features focus on:

1. **Volatility & stability** of employment
2. **Market concentration & dominance**
3. **Trend shape and curvature**
4. **Employment pressure on available space**
5. **Cycle & momentum in employment**
6. **Cross-property demand interaction**
7. **Regional spillovers**
8. **Shock detection & cumulative effects**

All of these are **used in top industrial real estate analytics**, especially for forecasting vacancy in multi-market, quarterly datasets like Moody's.

If you want, I can **build a table with 15-20 of these features, including formulas and short Python-ready logic**, so you can plug them directly into your modeling workflow.

Do you want me to do that?