

# 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 `trade_transportation_utilities_employment` and response `vacancy`. Help me understand how it worked

Got it! Let's break this down carefully so you can understand how `trade_transportation_utilities_employment` might affect `vacancy` in your real estate dataset. I'll cover the analysis step by step:

## 1. Understanding the Feature and Response

- **Feature: `trade_transportation_utilities_employment`**
  - This represents the total number of people employed in the **trade, transportation, and utilities sector** in a given market and quarter.
  - These sectors are closely linked to economic activity, population movement, and business operations in a city or region.
  - High employment in this sector usually signals:
    - Strong commerce (trade)
    - Active logistics and transport
    - Stable utility services (power, water, etc.)
- **Response: `vacancy`**
  - Vacancy is usually the proportion of unoccupied real estate (could be office, retail, industrial, etc.).
  - Low vacancy → high demand / tight market.
  - High vacancy → oversupply / weak demand.

## 2. Expected Relationship

In real estate economics, there are some intuitive relationships:

### 1. Trade / Retail:

- Higher employment in trade suggests more shopping activity → higher demand for retail spaces → **lower vacancy** in retail properties.

### 2. Transportation / Warehousing:

- More transport/logistics jobs → more industrial/warehouse demand → **lower vacancy** in industrial properties.

### 3. Utilities:

- Stable utilities employment indicates good infrastructure → indirectly supports commercial occupancy → **lower vacancy**.

✅ So generally, we expect a negative correlation between **trade\_transportation\_utilities\_employment** and **vacancy** in most real estate types.

## 3. How It Works in Time Series / Quarterly Data

Since your dataset is **quarterly per market**, some extra dynamics matter:

- **Lag Effects:**  
New hires or job growth might take a few quarters to affect real estate demand. You might see vacancy respond **with a lag** (1–2 quarters later).
- **Market Size Differences:**  
Absolute employment numbers might vary widely by market. Normalizing (employment per 1,000 people or per square foot of real estate) could improve predictive power.
- **Seasonality / Cycles:**  
Real estate vacancy can have quarterly or yearly cycles. You might need to detrend or seasonally adjust both employment and vacancy.

## 4. Feature Engineering Ideas

To better capture the effect on vacancy, you can derive some realistic features from **trade\_transportation\_utilities\_employment**:

### 1. Quarter-over-Quarter Growth:

```
python

emp_growth = current_quarter - previous_quarter
```

Captures acceleration/deceleration in sector employment.

### 2. Year-over-Year Growth:

Helps smooth seasonal effects.

### 3. Lagged Features:

- 1-quarter lag: **emp\_lag1**
- 2-quarter lag: **emp\_lag2**
- Captures delayed effect on vacancy.

### 4. Ratio to Total Employment:

```
python

ratio = trade_transportation_utilities_employment / total_non_agricultural_employment
```

Normalizes for market size.

## 5. Rolling Average (Trend):

- 4-quarter rolling average → smoother trend signal.

## 6. Interaction with Other Features:

- Example: `emp_growth * industrial_space_supply` → more predictive of vacancy changes in industrial properties.

# 5. How Top Real Estate Analysts Look at It

- They rarely use the raw employment number alone.
- They focus on **growth trends, ratios, and lags**, because real estate responds slowly to employment changes.
- They segment by property type:
  - Retail → trade employment
  - Industrial → transportation/logistics employment
  - Office → utilities / professional services employment

# 6. Quick Statistical Check

To see how it worked historically:

```
python

# Correlation
df[['trade_transportation_utilities_employment', 'vacancy']].corr()

# Lagged correlation
df['emp_lag1'] = df['trade_transportation_utilities_employment'].shift(1)
df[['emp_lag1', 'vacancy']].corr()
```

- Expect a **negative correlation** for most property types.
- Lagged features often improve correlation.

If you want, I can **create a table showing exactly how this feature historically affected vacancy for each property type in your dataset** — including lag effects and correlations — so you can see it in numbers and trends.

Do you want me to do that?



Help with out of box features engineering for this column.

Perfect — let's think like a **top real estate analyst** and create some **"out-of-the-box" features** from `trade_transportation_utilities_employment` that go beyond standard growth and lag features. Since you

are working with quarterly data and vacancy response, we want features that **capture trends, interactions, and market dynamics**. Here's a structured approach:

## 1. Advanced Lag & Momentum Features

- **Multi-Quarter Momentum:**

Capture the acceleration/deceleration in employment changes:

```
python

emp_momentum_2q = emp_q - emp_q_minus_2
emp_momentum_4q = emp_q - emp_q_minus_4 # year-over-year change
```

- **Cumulative Growth:**

Sum growth over last 4 quarters to capture sustained hiring trends:

```
python

emp_cum_growth = emp_q + emp_q_minus_1 + emp_q_minus_2 + emp_q_minus_3
```

## 2. Employment Concentration / Market Share Features

- **Share of Total Employment in Market:**

```
python

emp_ratio = trade_transportation_utilities_employment / total_non_agricultural_employment
```

- **Sector Diversity Index:**

How balanced is employment across sectors? High concentration in one sector may impact vacancy differently.

```
python

diversity_index = -sum((emp_sector_i / total_emp) * log(emp_sector_i / total_emp))
```

## 3. Interaction Features

- **Employment × Vacancy Lag Interaction:**

To model delayed impact of job growth on vacancy:

```
python

emp_lag1 * vacancy_lag1
```

- **Employment × Supply / Completions:**

Reflects if employment growth outpaces new construction:

```
python

emp_to_completions_ratio = emp_q / new_completions_q
```

## 4. Trend & Cyclical Features

- **Rolling Mean / Rolling Std:**

Captures trends and volatility over last 4 quarters:

```
python

emp_rolling_mean_4q
emp_rolling_std_4q
```

- **Quarterly Growth Acceleration:**

```
python

emp_growth_acceleration = (emp_q - emp_q_minus_1) - (emp_q_minus_1 - emp_q_minus_2)
```

- **Seasonal Index:**

Some markets have seasonal hiring in retail/transportation:

```
python

emp_seasonal_index = emp_q / mean(emp_q_same_quarter_prev_year)
```

## 5. Out-of-the-Box Real Estate Features

- **Employment Elasticity Feature:**

Measures sensitivity of vacancy to changes in employment (use past data):

```
python

emp_elasticity = (vacancy_q - vacancy_q_minus_1) / (emp_q - emp_q_minus_1)
```

- **Employment Volatility Feature:**

Volatile employment can indicate market instability → higher vacancy risk:

```
python

emp_volatility = std(emp_last_4_quarters)
```

- **Cross-Market Comparison:**

How does the market's employment growth compare with national / regional average?

```
python

emp_relative_growth = emp_q / emp_avg_region_q
```

## 6. Signals for Property-Type-Specific Vacancy

- **Retail Vacancy Predictor:**  $\text{trade\_emp\_growth\_ratio} = \text{trade\_emp} / \text{total\_trade\_emp\_region}$
- **Industrial Vacancy Predictor:**  $\text{transportation\_emp\_acceleration}$
- **Office Vacancy Predictor:**  $\text{utilities\_emp\_rolling\_mean}$

💡 **Tip:** Instead of blindly adding all features, you can use **feature importance ranking or causal analysis** to select the ones that actually move vacancy.

If you want, I can **create a ready-to-use list of 15–20 engineered features** from `trade_transportation_utilities_employment` specifically tailored for **quarterly real estate vacancy modeling**, which you can directly plug into your dataset.

Do you want me to do that next?



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

Ah, got it — you want **cutting-edge, “industry-grade” feature engineering** for `trade_transportation_utilities_employment` like top real estate analysts at Moody’s or CoStar would use. Let’s go beyond basic lags and rolling averages and focus on **features that capture market dynamics, elasticity, saturation, and leading indicators**. Here’s a curated set of **realistic, high-value features**:

## 1. Elasticity & Sensitivity Features

Top analysts often model **how vacancy reacts to employment changes**, not just absolute levels.

### 1. Vacancy-Employment Elasticity

- Measures how sensitive vacancy is to changes in employment:

```
python

emp_vacancy_elasticity = (vacancy_q - vacancy_q_minus_1) / (emp_q - emp_q_minus_1)
```

- Captures whether a small increase in jobs reduces vacancy significantly.

### 2. Employment Growth Acceleration

- Signals whether hiring is picking up faster than before:

```
python

emp_acceleration = (emp_q - emp_q_minus_1) - (emp_q_minus_1 - emp_q_minus_2)
```

- Strong positive acceleration often precedes declining vacancy.

## 2. Market Saturation & Supply Pressure

Top models consider **how employment interacts with existing inventory and completions**.

### 1. Employment per Square Foot of Inventory

- Normalizes employment relative to available space:

```
python

emp_per_sqft = trade_transportation_utilities_employment / total_available_space
```

### 2. Employment-to-Completion Ratio

- Detects if new hires are outpacing construction:

```
python

emp_to_new_completions = emp_q / completions_q
```

- High ratio → tighter market → lower vacancy risk.

## 3. Cross-Sector Interactions

Vacancy is rarely driven by one sector alone. Top analysts combine signals:

### 1. Interaction with Industrial/Commercial Employment

- Example: Transportation employment × Industrial employment

```
python

emp_transport_industrial = transportation_emp * industrial_emp
```

- Useful for predicting industrial and logistics space demand.

### 2. Relative Sector Share

- How dominant is trade/transport/utilities in total employment?

```
python

sector_share = trade_transport_utilities_emp / total_non_agricultural_emp
```

## 4. Trend & Momentum Features

Industry experts often use **trend, volatility, and rolling statistics**:

### 1. 4-Quarter Rolling Growth / Std Dev

- Captures momentum and employment stability:

```
python

emp_rolling_mean = rolling_mean(emp_last_4_quarters)
emp_rolling_std = rolling_std(emp_last_4_quarters)
```

## 2. Year-over-Year Seasonal Adjustment

- Removes seasonality effects in trade/retail hiring:

```
python

emp_seasonal_index = emp_q / mean(emp_q_same_quarter_last_year)
```

# 5. Leading & Lagging Indicators

Top real estate models exploit **lagged effects**, since vacancy responds slowly:

## 1. Lagged Employment Changes

- 1–4 quarter lags to model delayed vacancy impact:

```
python

emp_lag1, emp_lag2, emp_lag4
```

## 2. Lagged Vacancy × Employment Interaction

- Detects “market stickiness”:

```
python

emp_lag1 * vacancy_lag1
```

# 6. Volatility & Stability Features

High volatility in employment often signals **market risk**, which can increase vacancy:

## 1. Quarterly Employment Volatility

```
python

emp_volatility = std(emp_last_4_quarters)
```

## 2. Acceleration/Deceleration Flags

- Binary features: “Rapid growth” or “Rapid decline” in employment

```
python

emp_accel_flag = 1 if emp_acceleration > threshold else 0
```



## 7. Relative / Cross-Market Comparison

Analysts benchmark a market against peers:

### 1. Employment Relative to Regional Average

```
python

emp_relative = emp_q / regional_avg_emp_q
```

### 2. Percentile Rank Across Markets

- Rank a market's employment growth in the national or regional distribution:

```
python

emp_percentile = rank(emp_growth_q) / total_markets
```

### ✓ Why these features work in top real estate models:

- Capture **trends, momentum, and delayed effects** (vacancy responds slowly to employment changes).
- Normalize for **market size and supply constraints**.
- Combine **sector interactions** to predict property-type-specific vacancies.
- Include **volatility & elasticity**, which provide early warning signals for oversupply or undersupply.

If you want, I can **draft a ready-to-use table of 20+ engineered features from trade\_transportation\_utilities\_employment** with exact formulas and recommended lags, ready to plug into your Moody's dataset for quarterly vacancy modeling.

Do you want me to do that?



Show me some more features engineering with trade\_transportation\_utilities\_employment which are realistic and don't try to keep the earlier feature.

Perfect — let's get creative with **fresh, realistic feature engineering ideas** for trade\_transportation\_utilities\_employment that are **different from the standard lags, growth, ratios, or rolling averages** we discussed before. I'll focus on **top real estate-industry-style signals**.

## 1. Structural / Distribution Features

These capture **how employment is distributed or concentrated over time or sectors**:

### 1. Quarterly Employment Skew

- Measures whether employment growth is concentrated in a few sub-sectors:

```
python

emp_skew = skew([trade_emp, transport_emp, utilities_emp])
```

- High skew → uneven sector growth → may affect specific property types.

## 2. Quarterly Employment Kurtosis

- Detects extreme spikes in hiring:

```
python

emp_kurtosis = kurtosis([trade_emp, transport_emp, utilities_emp])
```

- Could signal rapid sector expansion or contraction → early vacancy signal.

## 3. Share of Rapidly Growing Sub-Sectors

- Identify which part of the sector is accelerating:

```
python

fast_growth_ratio = sum([emp_growth_i > threshold]) / 3
```

- Example: if 2 out of 3 sub-sectors are growing rapidly, flag = 0.66.

# 2. Market Sentiment / Pressure Indicators

These features measure **market stress or opportunity** via employment patterns:

## 1. Employment Saturation Index

- Employment growth relative to historical max:

```
python

saturation_index = emp_q / max(emp_last_8_quarters)
```

- Near 1 → market may be saturated → vacancy could rise if supply increases.

## 2. Employment Pressure Score

- Measures pressure on local real estate:

```
python

emp_pressure = emp_q / (vacancy_q + 1) # avoid division by zero
```

- High value → high jobs, low vacancy → strong demand signal.

## 3. Sector Volatility Spread

- Difference between most and least volatile sub-sectors:

```
python
```

```
volatility_spread = max(std([trade, transport, utilities])) - min(std([trade, transport, utilities]))
```

- Indicates instability in the sector composition.

## 3. Leading / Coincident Indicators

Top analysts like features that **predict vacancy before it moves**:

### 1. Employment Trend Break

- Detect sudden reversals in growth:

```
python
```

```
trend_break = 1 if (emp_q - emp_q_minus_1) * (emp_q_minus_1 - emp_q_minus_2) < 0 else 0
```

- Signals inflection points in hiring → vacancy may lag.

### 2. Relative Growth to Long-Term Average

```
python
```

```
emp_vs_longterm = emp_q / mean(emp_last_12_quarters)
```

- Captures whether current hiring is unusually high/low.

### 3. Employment Concentration Index (Gini)

- Measures inequality among sub-sectors:

```
python
```

```
gini_index = gini([trade_emp, transport_emp, utilities_emp])
```

- Uneven employment → may favor some property types over others.

## 4. Interaction & Nonlinear Features

Real estate models often include **interaction or non-linear relationships**:

### 1. Squared Employment Growth

```
python
```

```
emp_growth_squared = (emp_q - emp_q_minus_1)**2
```

- Captures strong spikes in hiring → non-linear vacancy response.

### 2. Employment × Market Size

```
python

emp_market_interaction = emp_q * total_population_q
```

- Larger cities with same employment growth may have different vacancy dynamics.

### 3. Employment × Inflation / Rent Index

- If you have CPI or rent data:

```
python

emp_rent_interaction = emp_q * rent_index_q
```

- Captures affordability-driven demand effects.

## 5. Structural Break / Regime Features

- Detect **periods of unusual employment growth or slowdown**:

### 1. Quarterly Growth Outlier Flag

```
python

emp_outlier_flag = 1 if emp_growth_q > mean(emp_growth_last_8q) +
2*std(emp_growth_last_8q) else 0
```

### 2. Rolling Max/Min Differences

```
python

emp_rolling_diff = rolling_max(emp_last_8q) - rolling_min(emp_last_8q)
```

- Useful to spot extreme cycles that may affect vacancy.

### Why these features are “realistic top-industry style”:

- They go **beyond lags, ratios, and simple rolling averages**.
- Capture **distribution, volatility, regime shifts, nonlinear effects**.
- Can help models **predict vacancy by property type or market more precisely**.

If you want, I can **create a full ready-to-use list of 15–20 brand-new engineered features from this employment column**, with **formulas and example code**, ready for your Moody’s quarterly dataset.

Do you want me to do that next?