

Retail sales and vacancy rate



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 `Total_Retail_Sales` and response `vacancy_rate`. Help me understand how it works

Got it, let's break this down carefully. You're looking at **Total_Retail_Sales** as a feature and **vacancy_rate** as the response, in a **quarterly real estate dataset** across multiple markets. Here's how we can understand the relationship:

1. What the feature represents

- **Total_Retail_Sales:** This represents the total value of retail sales in a market for a given quarter.
 - High values → strong consumer spending → potentially strong demand for retail space.
 - Low values → weaker consumer spending → lower demand for retail space.
- **Vacancy_rate:** Percentage of available retail space that is unoccupied.
 - High vacancy → less demand or oversupply.
 - Low vacancy → strong demand and good occupancy.

2. Intuition of the relationship

- Retail sales are usually **negatively correlated** with vacancy rate.
 - If retail sales increase, retailers are more likely to occupy spaces → vacancy decreases.
 - If retail sales fall, retailers may close stores or delay expansion → vacancy increases.
- **Lagged effects** often appear:
 - Retailers may react to past quarters' sales trends.
 - Example: Retail sales in Q1 might affect vacancy rate in Q2 (due to lease negotiations, openings, or closures).

3. Ways to feature engineer Total_Retail_Sales

To make it more useful for predicting **vacancy_rate**, real estate professionals often create derived features like:

1. Growth metrics

- Quarter-over-quarter growth:

$$\text{Retail_Sales_Growth} = \frac{\text{Sales}_{\{t\}} - \text{Sales}_{\{t-1\}}}{\text{Sales}_{\{t-1\}}}$$

- Year-over-year growth: compares same quarter last year (good for seasonal adjustments).

2. Rolling averages

- 2-quarter or 4-quarter rolling averages smooth volatility:

$$\text{Rolling_Avg} = \frac{\text{Sales}_t + \text{Sales}_{t-1} + \dots + \text{Sales}_{t-n+1}}{n}$$

3. Lag features

- Lagged values of sales (t-1, t-2, ...) capture delayed effects on vacancy.
 - Example: Q1 sales might affect Q2 vacancy → include `Sales_Lag_1Q`.

4. Market-relative metrics

- Normalize by market size or population: sales per capita.
- Compare current sales to historical market trend (z-score or % deviation).

5. Seasonality indicators

- Retail has strong seasonality (holidays, back-to-school, etc.).
- Features like `Q4_sales` or `Holiday_season_sales` can help model seasonal vacancy changes.

4. Advanced perspectives

- **Elasticity of vacancy vs. retail sales:**
 - You can estimate how sensitive vacancy is to sales changes.
 - Example: a 10% drop in sales might increase vacancy by 0.5%—helpful for scenario analysis.
- **Cross-market interaction:**
 - Retail sales in one market may slightly affect nearby submarkets (retail spillover effect).
- **Segmentation by property type:**
 - If your dataset has property types (malls, street retail, big-box), the effect of retail sales on vacancy may differ by type.

✓ Summary:

`Total_Retail_Sales` is a **leading indicator of retail space demand**, but its effect on vacancy is usually **delayed and smoothed**. Feature engineering like growth rates, lags, rolling averages, and seasonality adjustments will make it much more predictive.

If you want, I can create a **list of 10+ top-notch feature engineering ideas for `Total_Retail_Sales`**, exactly like Moody's/real estate pros use, ready to plug into your models.

Do you want me to do that?



Help with out of box features engineering for this column.

Perfect! Let's go beyond the standard lag and rolling average features and focus on **out-of-the-box, top-tier feature engineering ideas** for `Total_Retail_Sales` in real estate vacancy prediction. These are

techniques that advanced analysts in the real estate and financial sectors use to extract more predictive power.

1. Relative Performance Features

- **Sales vs. Historical Trend:**

Compare current quarter sales to a multi-year trend for the same quarter:

$$\text{Sales_vs_Trend} = \frac{\text{Sales}_t - \text{Avg}(\text{Sales}_{t-4}, t-8, t-12))}{\text{Std}(\text{Sales}_{t-4}, t-8, t-12))}$$

Captures whether the market is over- or under-performing relative to its seasonal pattern.

- **Sales Share of Market:**

Normalize by total sales across all markets for the quarter:

$$\text{Market_Share} = \frac{\text{Sales}_{\text{market},t}}{\sum_{\text{all_markets}} \text{Sales}_t}$$

Useful for multi-market models.

2. Change & Momentum Features

- **Acceleration of Sales:**

Not just growth, but the rate of growth change:

$$\text{Acceleration} = \text{Sales}_t - 2 \cdot \text{Sales}_{t-1} + \text{Sales}_{t-2}$$

Detects sudden market slowdowns or surges before vacancy reacts.

- **Volatility / Risk Features:**

Rolling standard deviation of sales over last 4 quarters:

$$\text{Sales_Volatility} = \text{Std}(\text{Sales}_{t-3:t})$$

High volatility can indicate uncertain market conditions → higher vacancy risk.

3. Seasonality & Cyclical Features

- **Seasonal Index:**

Ratio of current quarter sales to average for that same quarter historically:

$$\text{Seasonality_Factor} = \frac{\text{Sales}_t}{\text{Avg}(\text{Sales}_{t-4,t-8,t-12})}$$

Highlights quarters where sales are unusually high/low.

- **Quarterly Cycles:**

Encode Q1-Q4 as cyclic features using sine/cosine transforms for smooth seasonality handling.

4. Cross-Feature Interactions

- **Sales per Capita / per Household:**

$$Sales_per_Household = \frac{Total_Retail_Sales}{Total_Households}$$

Reflects real market demand more accurately than raw sales.

- **Sales-to-Vacancy Lag Ratio:**

Create a feature like:

$$Sales_Lag1/Vacancy_Lag1$$

Captures how past sales movements relate to vacancy trends.

- **Retail Sales vs. Labor Force:**

Interaction term:

$$Retail_Sales \times Total_Labor_Force$$

Helps capture the economic environment effect.

5. Trend & Signal Extraction

- **Rolling Z-Scores:**

Standardize sales over rolling window (e.g., last 4 quarters):

$$Z = \frac{Sales_t - \text{mean}_{\text{window}}}{\text{std}_{\text{window}}}$$

Detects unusual surges or drops.

- **Cumulative Sales Growth:**

Sum growth over past N quarters to detect persistent trends rather than single-quarter spikes.

- **Retail Sales Momentum Signal:**

Encode positive/negative momentum as categorical:

- +1 = strong increase over past 2 quarters
- 0 = stable
- -1 = decline

6. Market Health Indicators

- **Retail Sales Shock:**

Flag quarters with >X% deviation from historical average. Useful as a stress indicator for vacancy prediction.

- **Quarterly Retail-to-GDP Ratio (if available):**

Captures relative importance of retail activity in the overall economy.

- **Retail Concentration Index:**

If submarkets are available, compute Gini or Herfindahl index of sales across submarkets. Uneven sales can indicate risky areas prone to vacancies.

7. Advanced Time-Series Features

- **Exponential Weighted Moving Average (EWMA)**

Captures recent trends more heavily than older data:

$$EWMA_t = \alpha \cdot Sales_t + (1 - \alpha) \cdot EWMA_{t-1}$$

- **Sales Trend Slope**

Fit a linear trend on last N quarters to quantify increasing/decreasing momentum numerically.

- **Fourier Transforms / Seasonal Decomposition**

Decompose sales into trend + seasonal + residual → feed components separately to model.

✓ Summary of Out-of-the-Box Ideas

- Relative performance (trend vs. historical, market share)
- Acceleration & volatility metrics
- Seasonality encoding (cyclic, seasonal index)
- Cross-feature interactions (per capita, labor, lag ratios)
- Signal extraction (z-scores, momentum)
- Market health indicators (sales shock, concentration)
- Advanced time-series features (EWMA, trend slope, Fourier)

If you want, I can **make a concrete table of 15–20 engineered features for Total_Retail_Sales ready to plug into your Moody's dataset**, with formulas, descriptions, and type (numeric/categorical), exactly like top real estate analysts do.

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 **truly top-tier, industry-level feature engineering for Total_Retail_Sales** that goes beyond standard lag, rolling, or seasonality features—stuff that **real estate analysts at firms like Moody's, CBRE, or JLL actually use to forecast vacancy rates**. Let's go deep.

1. Demand-Supply Alignment Features

- **Retail Sales per Square Foot of Available Retail Space**

$$Sales_per_SqFt = \frac{Total_Retail_Sales}{Total_Retail_Space_Occupied + Vacant}$$

Insight: Shows whether the retail sales in a market are enough to support current retail supply. Low values → risk of vacancy growth.

- **Retail Sales vs. New Retail Construction**

$$Sales_to_New_Construction = \frac{Total_Retail_Sales}{New_Retail_Inventory}$$

Insight: Detects oversupply in markets, a major vacancy driver.

- **Retail Sales per Tenant / Anchor Ratio**

If data on major tenants exists, normalize sales per major tenant or anchor store. Identifies markets dependent on few big retailers—more sensitive to closures.

2. Market Momentum & Shock Features

- **Shock Detection (Outlier Quarters)**

Identify sales spikes/drops beyond 2–3 standard deviations from historical trends. Encode as binary or categorical “shock” features.

Insight: Retail shocks often precede vacancy changes.

- **Persistence of Sales Growth**

Count consecutive quarters of growth or decline.

Insight: Vacancy reacts more to sustained trends than one-off changes.

- **Sales Trend Acceleration**

Difference of differences over last 3–4 quarters:

$$Acceleration = (Sales_t - Sales_{t-1}) - (Sales_{t-1} - Sales_{t-2})$$

Insight: Early warning for sudden retail expansion or contraction.

3. Cross-Market & Competitive Features

- **Market Share Relative to Nearby Markets**

Normalize each market’s sales by the sum of nearby submarkets. Captures competitive pressures.

- **Retail Diversification Index**

Use sales distribution across retail segments (groceries, apparel, electronics). Compute **Herfindahl Index** or Gini coefficient.

Insight: Highly concentrated markets are riskier—vacancy rises faster if a segment declines.

- **Cannibalization Ratio**

If new retail openings are in same market/submarket:

$$Cannibalization = Newstore_sales / Total_Retail_sales$$

Captures internal competition causing vacancy.

4. Macro & Socioeconomic Alignment

- **Sales vs. Labor Force or Population**

$$Retail_Intensity = \frac{Total_Retail_Sales}{Total_Labor_Force} \quad \text{or} \quad /Population$$

Insight: Indicates whether sales are supported by population/economic base.

- **Sales to Disposable Income Ratio**

If income data is available:

$$Retail_Spending_Rate = \frac{Total_Retail_Sales}{Disposable_Income}$$

Helps capture overspending/underspending trends that affect retail demand.

- **Employment & Sector-Specific Sales Interaction**

Combine with employment in retail/hospitality sectors:

$$Sales \times Retail_Employment$$

Signals economic health at sector level.

5. Seasonal & Event-Specific Features

- **Holiday/Seasonal Sales Index**

Normalize each quarter by historic seasonal trend (Q4 spikes, back-to-school, Black Friday).

- Ratio to 3-year average same quarter: identifies over/underperformance.

- **Event-Driven Features**

Encode major local events affecting retail: openings of malls, closures of anchors, local policy changes.

Insight: Vacancy is often driven by one-off shocks in retail-heavy markets.

6. Advanced Time-Series & Statistical Features

- **Exponential Weighted Moving Average (EWMA) of Sales & Growth**

Prioritizes recent performance for vacancy forecasting.

- **Fourier or Wavelet Decomposition**

Separates trend, seasonal, and high-frequency variations. Use components as separate predictive features.

- **Rolling Quantiles / Percentiles**

Identify where current sales fall within historical distribution (e.g., 90th percentile spike → low vacancy likely).

- **Autoregressive Residual Features**

Fit AR model on historical sales → use residuals as predictive signal (unexpected changes often drive vacancy shifts).

✓ Summary of Top-Notch Features for Total_Retail_Sales

Category	Example Features	Insights
Demand-Supply	Sales/SqFt, Sales/New Construction	Detect oversupply & occupancy risk
Momentum/Shock	Sales Shock, Acceleration, Growth Persistence	Early warning for vacancy trend
Cross-Market	Market Share, Diversification Index, Cannibalization	Competitive risk & concentration effects
Macro Alignment	Sales/Population, Sales/Labor, Sales/Income	Local economic support for retail demand
Seasonal/Event	Holiday Sales Index, Event-driven	Captures predictable & one-off seasonal effects
Time-Series	EWMA, Fourier components, Rolling Quantiles, AR Residuals	Advanced trend & anomaly detection

If you want, I can **create a ready-to-use feature table of 20-25 engineered features for Total_Retail_Sales**, with **formulas, descriptions, type (numeric/categorical), and implementation hints**, exactly like a Moody's data science team would.

Do you want me to do that next?



So me some more features engineering with Total_Retail_Sales.

Absolutely! Let's go **even deeper into advanced, high-value feature engineering for Total_Retail_Sales** in real estate vacancy modeling. I'll focus on ideas that top analysts actually use in practice, going beyond the usual lags, rolling averages, or simple growth rates.

1. Advanced Trend & Momentum Features

- **Cumulative Growth Over N Quarters**

Sum up quarter-over-quarter growth over past 2-4 quarters to capture sustained trends:

$$Cumulative_Growth = \sum_{i=1}^N \frac{Sales_{t-i+1} - Sales_{t-i}}{Sales_{t-i}}$$

- **Momentum Category**

Encode sales trend into categories:

- +2 = accelerating growth
- +1 = stable growth
- 0 = flat
- -1 = decline
- -2 = accelerating decline

Insight: Easier for tree-based models to capture non-linear trends.

- **Slope of Linear Trend**

Fit a linear regression over the last 4–8 quarters of sales and use the slope as a feature. Captures early signals of increasing/decreasing demand.

2. Market Stress / Anomaly Features

- **Z-Score of Current Sales**

Measure deviation from historical mean over 8–12 quarters:

$$Z = \frac{Sales_t - \mu_{past}}{\sigma_{past}}$$

Extreme positive/negative values indicate unusual market conditions.

- **Sales Shock Flags**

Binary feature if current quarter sales $> \pm 2$ standard deviations from last year's same quarter. Predicts potential future vacancy changes.

- **Rolling Sales Volatility**

Standard deviation over past 4–8 quarters. High volatility = uncertain market → potential vacancy risk.

3. Relative / Normalized Features

- **Sales per Capita / Household / Worker**

$$Sales_per_Capita = \frac{Total_Retail_Sales}{Population}$$

Adjusts for market size; small markets may have high sales but low absolute demand.

- **Sales Share Within City or Region**

$$Market_Share = \frac{Sales_{market}}{\sum_{markets_in_region} Sales}$$

Captures competitive position of a market.

- **Sales Relative to Historical Peak**

Ratio of current sales to the highest historical sales in the market. Detects markets in decline vs. markets near capacity.

4. Interaction Features

- **Sales × Retail Employment**

$$Sales \times Retail_Employment$$

Captures economic support for retail demand in that market.

- **Sales × Vacancy Lag**
Interaction of lagged vacancy and current sales: identifies if strong past vacancy moderates effect of sales on current vacancy.
- **Sales × New Supply / Construction**
Ratio of sales to new inventory indicates market absorption capacity.

5. Seasonality & Event Features

- **Seasonal Index**
Compare sales to the same quarter in previous years:

$$Seasonal_Index = \frac{Sales_t}{Average_Sales_{same_quarter_last_3_years}}$$

Highlights whether sales are unusually strong or weak for that season.

- **Holiday / Event Adjustments**
Encode quarters with major events (e.g., holiday sales spikes, major mall openings/closures) as categorical or binary features.
- **Quarter Cosine/Sine Encoding**
Transform quarter (Q1-Q4) into cyclic features for smooth seasonal modeling:

$$Quarter_Sin = \sin\left(\frac{2\pi Q}{4}\right), \quad Quarter_Cos = \cos\left(\frac{2\pi Q}{4}\right)$$

6. Advanced Time-Series Features

- **EWMA (Exponential Weighted Moving Average)**

$$EWMA_t = \alpha \cdot Sales_t + (1 - \alpha) \cdot EWMA_{t-1}$$

Prioritizes recent trends.

- **AR Residuals**
Fit an AR(1) or AR(2) model on past sales; use residuals as a feature. Captures unexpected changes in sales not explained by trend.
- **Fourier / Seasonal Decomposition Components**
Decompose sales into trend + seasonal + residual; feed each as a separate feature.

7. “Top-Notch” Real Estate Features

- **Retail Sales per Available SqFt**
Indicates if retail demand is sufficient for the existing supply.
- **Sales-to-Rent Ratio (if rental data available)**
Shows affordability of retail space relative to sales.

- **Retail Concentration Index**

Compute Gini or Herfindahl Index over retail categories or tenants; markets with highly concentrated sales are more sensitive to shocks → potential vacancy risk.

- **Lagged Growth Ratios vs. Neighboring Markets**

Compare growth in your market to nearby markets over last 2–4 quarters. Detects competitive pressure and relative market strength.

If you want, I can **create a concrete “plug-and-play” feature table for Total_Retail_Sales**, with **20–25 features, formulas, type (numeric/categorical), and explanation of how each helps predict vacancy**, just like Moody’s data science teams would implement.

Do you want me to do that next?