

GMP and vacancy relationship



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

Perfect — you're dealing with a key **macroeconomic driver** here. Let's break this down step-by-step so you can fully understand how **Gross Metro Product (GMP)** interacts with **vacancy rate** in a real estate dataset like Moody's.

1. What is Gross Metro Product (GMP)?

Gross Metro Product (GMP) — sometimes called **Gross Metropolitan Product** — is the **total value of goods and services** produced within a metropolitan area (a city and its surrounding region).

Think of it as **GDP at the metro level**.

- It measures **economic output**.
- Higher GMP = a larger or faster-growing economy in that metro.
- It reflects the strength of business activity, consumer spending, and employment.

So when Moody's reports `Gross_Metro_Product_Total`, it's basically tracking **how economically strong** each city (market) is over time — quarterly.

2. How GMP connects to Real Estate Vacancy Rate

Now, let's link this to **vacancy** (your response variable).

a. Direct Economic Mechanism

When a metro's economy (GMP) grows:

- **Businesses expand**, increasing demand for **office and industrial** space.
- **People get employed**, driving demand for **residential and retail** properties.
- → Vacancy rates **decline** (more space gets filled).

When GMP **contracts** (falls or slows):

- Firms **reduce hiring or close offices**.
- Consumer activity slows → less retail demand.
- → Vacancy rates **rise**.

So generally:

- ▲ GMP ↑ → Vacancy ↓
- ▼ GMP ↓ → Vacancy ↑

But the **strength and lag** of this relationship depends on the property type and market.

3. Why the Relationship Isn't Instantaneous

Vacancy rates **lag** behind economic activity:

- Businesses don't move immediately after GDP shifts.
- New construction can temporarily increase vacancy even when the economy is growing.

Hence, analysts often use **lagged features** like:

- GMP_{t-1} (previous quarter's GMP)
 - ΔGMP_{t-1} (quarter-over-quarter growth)
 - Or even a **rolling average** to smooth volatility.
-

4. Typical Feature Engineering Approaches (used by top firms)

You can make **derived features** from GMP that capture economic momentum and inflection points. Here are **industry-grade transformations**:

(A) Growth and Acceleration

1. Quarter-over-Quarter Growth Rate

$\text{GMP_Growth_QoQ} = \frac{GMP_t - GMP_{t-1}}{GMP_{t-1}}$
→ captures short-term economic momentum.

2. Year-over-Year Growth Rate

$\text{GMP_Growth_YoY} = \frac{GMP_t - GMP_{t-4}}{GMP_{t-4}}$
→ accounts for seasonality and long-term trends.

3. Acceleration (2nd derivative)

Difference in growth rates:

$\text{GMP_Acceleration} = \text{GMP_Growth_QoQ}_t - \text{GMP_Growth_QoQ}_{t-1}$
→ positive acceleration = improving economy, likely future lower vacancy.

(B) Smoothing and Momentum Indicators

4. Rolling Mean (3- or 4-quarter)

Captures local economic trend:

$\text{GMP_MA_4Q} = \text{mean}(GMP_{t-3:t})$

5. Rolling Growth Momentum

$\text{GMP_Momentum} = \frac{GMP_t - GMP_{t-4}}{GMP_{t-4}}$

→ smooths out short-term noise, highlights sustained growth periods.

(C) Elasticity or Relative Measures

6. GMP per Capita

$$\text{GMP_Per_Capita} = \frac{\text{GMP}_t}{\text{Population}_t}$$

→ more meaningful for comparing metros of different sizes.

7. Vacancy Elasticity Proxy

You can model how much vacancy changes when GMP changes:

$$\text{Vacancy_Elasticity} = \frac{\% \Delta \text{Vacancy}}{\% \Delta \text{GMP}}$$

→ Useful in explanatory modeling to show sensitivity.

(D) Cross-Feature Interactions

8. Interaction with Employment

Combine GMP and employment to reflect productivity or output per job:

$$\text{Productivity_Index} = \frac{\text{GMP}_t}{\text{Total_Employment}_t}$$

→ higher productivity might signal automation, less space demand.

9. Interaction with Construction or Inventory

$$\text{GMP_to_New_Supply_Ratio} = \frac{\text{GMP}_{\text{Growth}}}{\text{New_Construction_Volume}}$$

→ shows if economic growth can absorb new space coming to market.

5. Interpreting in Modeling Context

When you model $\text{Vacancy} \sim \text{GMP}$:

- Expect a **negative coefficient** (higher GMP → lower vacancy).
- But check:
 - **Lag effects** (t-1 or t-2 might work better).
 - **Nonlinear effects** (log or polynomial terms may fit better).
 - **Interactions** (e.g., $\text{GMP} \times \text{Employment}$).

6. Example Real Estate Insights

Market	Observation	Implication
Austin	GMP growing + tech expansion	Office & residential vacancy likely declining
New York	GMP slowing + new office supply	Vacancy may rise despite large economy
Atlanta	Steady GMP + logistics growth	Industrial vacancy stays low

If you want, I can help you **design a feature engineering block specifically for GMP** — with code-ready transformations (like how you'd implement it in Python or SQL for your Moody's dataset).

Would you like me to do that next?



Help with out of box features engineering for this column.

Excellent — let's go beyond standard growth rates and create **"out-of-box" feature engineering ideas** for **Gross_Metro_Product_Total** (GMP) that top-tier real estate analytics and institutional investors (like Moody's, CBRE, and Blackstone) often use.

We'll focus on **economic-cycle intelligence, cross-domain signals, and predictive transformations** that enhance vacancy modeling.

Out-of-the-Box Feature Engineering Ideas for Gross_Metro_Product_Total

1. Economic Cycle Positioning (Phase Indicator)

Idea: Identify which part of the business cycle each metro is in — expansion, slowdown, contraction, or recovery.

How:

- Compute **GMP growth rate** (YoY or QoQ).
- Smooth using 4-quarter rolling average.
- Classify:
 - Growth > +2% → Expansion
 - 0–2% → Mild Growth
 - -2–0% → Slowdown
 - < -2% → Recession

Feature name: **GMP_Cycle_Phase** (categorical or ordinal)

Why: Vacancy behavior differs by cycle — during early expansion, vacancy drops sharply; during late expansion, it stabilizes.



2. Economic Stress Index (Volatility Measure)

Idea: Capture the *volatility* of GMP growth to represent how uncertain the local economy is.

How:

$\text{GMP_Volatility} = \text{std}(\text{GMP_Growth}_{t-3:t})$

Why: Markets with unstable GMP growth often show delayed or unstable real estate absorption → higher vacancy volatility.

3. Economic Momentum Divergence

Idea: Compare the metro's GMP growth to the **national average** GMP growth.

How:

$$\text{GMP_Relative_Growth} = \text{GMP_Growth_Metro} - \text{GMP_Growth_National}$$

Why: If a metro grows faster than the nation → outperforming market, likely lower vacancy.

Feature name: GMP_Alpha (analogy to financial alpha).

4. Turning Point Signal (Growth Direction Change)

Idea: Identify when GMP growth changes direction (inflection points).

How:

- Compute lagged growth: GMP_Growth_t and GMP_Growth_{t-1}.
- If $\text{sign}(\text{GMP_Growth}_t - \text{GMP_Growth}_{t-1})$ changes → turning point.
- Encode as binary: 1 if turning point, else 0.

Feature name: GMP_Turning_Point

Why: Vacancy rate usually reacts **1-2 quarters after** a turning point in economic activity.

5. Output-to-Vacancy Elasticity Proxy

Idea: Measure how vacancy responds to GMP changes *within each market*.

How:

$$\text{Elasticity_Index} = \frac{\Delta \text{GMP}}{\Delta \text{Vacancy}_{\text{lag1}}}$$

Compute at metro level using rolling windows.

Why: Indicates sensitivity — some markets (like Austin) have high responsiveness, others (like NYC) low responsiveness.

6. Economic Saturation Index

Idea: Estimate whether current GMP is above its long-term trend → an indicator of overheating or underperformance.

How:

- Compute **long-term rolling average (8-12 quarters)**.
- Compare current value:

$$\text{GMP_Saturation} = \frac{\text{GMP}_t - \text{GMP}_{\text{MA12}}}{\text{GMP}_{\text{MA12}}}$$

Why:

- Positive = economy above potential (vacancy may rise soon as market overheats).
- Negative = economy below potential (vacancy may decline later as recovery starts).

Idea: Use a **Hodrick–Prescott (HP) filter** or **rolling linear regression** to remove trend and isolate cyclical component.

How (conceptually):

$$GMP_{cycle} = GMP_t - Trend(GMP_t)$$

Why: The cyclical component of GMP aligns tightly with vacancy movements (macro-level correlation).

Feature name: GMP_Cycle_Component.

Idea: Combine GMP with sectoral employment or construction data.

Examples:

- $GMP_per_Construction_Worker = GMP / Construction_Employment$
- $GMP_per_Office_Worker = GMP / Office_Employment$

Why: Shows whether economic output is efficiently supported by workforce — strong productivity often correlates with lower vacancy in prime sectors.

Idea: Capture delayed relationships (since vacancy reacts slowly).

How:

Create lagged versions:

- GMP_t-1, GMP_t-2, GMP_t-3
- GMP_Growth_t-1, GMP_Growth_t-2

And their averages:

$$\text{Lagged_Avg} = \frac{GMP_t + GMP_{t-1} + GMP_{t-2}}{3}$$

Why: Vacancy might react to sustained economic trends rather than single-quarter shocks.

Combine GMP with other features in your Moody's dataset:

Interaction	Formula	Interpretation
GMP × Total_Employment	$GMP_t * Employment_t$	Aggregate economic capacity
GMP / SP500_Price_Index	—	Local vs national financial divergence

Interaction	Formula	Interpretation
$\Delta GMP \times \text{New_Construction}$	—	Can new supply keep up with growth?
$GMP_Growth \times \text{Inflation}$	—	Real growth pressure on occupancy

11. Relative Positioning Among Markets

Idea: Rank each metro's GMP growth each quarter relative to others.

How:

$\text{GMP_Rank_Percentile} = \text{percentile_rank}(GMP_Growth_t)$

Why: Relative economic strength often explains cross-market vacancy differences even better than absolute GMP.

12. Predictive Composite Feature: Economic Health Score

Combine multiple signals:

$EHS = 0.4 \times GMP_Growth_YoY + 0.3 \times (-GMP_Volatility) + 0.3 \times GMP_Relative_Growth$

→ Scaled 0-1 indicator of local economic health.

Why: Acts as a single meta-feature summarizing the market's economic strength, used by hedge funds and institutional forecasters.

Summary Table of Feature Ideas

#	Feature Name	Type	Core Idea	Expected Vacancy Relation
1	GMP_Cycle_Phase	Categorical	Business cycle stage	Expansion ↓ vacancy
2	GMP_Volatility	Continuous	Economic uncertainty	High volatility ↑ vacancy
3	GMP_Alpha	Continuous	Relative growth vs nation	Positive ↓ vacancy
4	GMP_Turning_Point	Binary	Economic inflection	Turning point → vacancy change soon
5	Elasticity_Index	Continuous	Sensitivity metric	High → faster vacancy response
6	GMP_Saturation	Continuous	Above/below trend	Above → future vacancy ↑
7	GMP_Cycle_Component	Continuous	Detrended cycle	Positive → vacancy ↓
8	Productivity_Proxy	Continuous	Output per worker	High → stable occupancy
9	Lagged_Avg	Continuous	Momentum	High → sustained lower vacancy
10	Interaction Terms	Continuous	Cross-factor insight	Depends on direction
11	GMP_Rank_Percentile	Continuous	Relative metro position	High rank → lower vacancy
12	Economic_Health_Score	Composite	Weighted indicator	High → lower vacancy

Would you like me to now show how to **implement these in Python (Pandas)** — including rolling, lag, and interaction code snippets — ready to apply on your Moody's dataset?



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 want **truly top-tier, “institutional-grade” feature engineering** that the best real estate analysts (like CBRE, JLL, Moody’s Analytics, or Blackstone) actually use for metro-level GMP data to predict vacancy. Let’s go **beyond standard growth rates and moving averages** and focus on **sophisticated, predictive, and cross-market features**.

🏆 Top-Notch Feature Engineering for Gross_Metro_Product_Total

1 Lagged Economic Stress & Recovery Indicators

- **Concept:** Vacancy responds with delay to economic cycles. Top firms create multi-quarter **lag features** to capture delayed effects.
- **Implementation ideas:**
 - $GMP_{t-1}, GMP_{t-2}, GMP_{t-3} \rightarrow$ short-term lag
 - $GMP_{rolling_avg_4Q} \rightarrow$ trend smoothing
 - $GMP_{growth_acceleration} = (GMP_t - GMP_{t-1}) - (GMP_{t-1} - GMP_{t-2}) \rightarrow$ measures sudden change in momentum
- **Why top-notch:** Captures “economic inertia” — key for predicting office and industrial vacancy.

2 GMP Cycle Decomposition

- **Concept:** Separate GMP into **trend vs. cyclical components**.
- **Methods used by top analysts:**
 - **Hodrick-Prescott (HP) filter** or **Baxter-King filter**
 - $GMP_{cycle_component} = GMP - Trend(GMP)$
- **Why:** Vacancy is more sensitive to cyclical deviations than absolute GDP.
- **Example:** If a metro is above its trend, it may experience oversupply, causing vacancy to rise.

3 Cross-Metro Comparative Features

- **Concept:** Measure a metro’s GMP relative to peer metros or national average.
- **Top metrics:**
 - $GMP_{relative_growth} = GMP_{growth_metro} - GMP_{growth_national}$
 - $GMP_{percentile_rank}$ among all metros
- **Why:** Captures local economic outperformance or underperformance — often predicts where vacancy will fall or rise faster than the national average.

4 Sectoral GMP Contribution Ratios

- **Concept:** GMP alone is useful, but top firms break it down by sectors to understand **which parts of the economy drive real estate demand**.
- **Examples:**
 - $\text{Tech_GMP} / \text{Total_GMP} \rightarrow$ high tech concentration may signal office demand volatility
 - $\text{Retail_GMP} / \text{Total_GMP} \rightarrow$ proxy for retail absorption
 - $\text{Industrial_GMP} / \text{Total_GMP} \rightarrow$ predicts warehouse vacancy trends
- **Why:** Helps link economic activity directly to property types.

5 GMP Elasticity & Sensitivity Metrics

- **Concept:** How sensitive is vacancy to GMP changes?
- **Implementation:**
 - Rolling elasticity: $\Delta \text{Vacancy} / \Delta \text{GMP}$ over 4–8 quarters
 - Weighted elasticity per property type (office, retail, industrial)
- **Why top-notch:** Directly quantifies “how much vacancy reacts to economic growth” — used for predictive stress-testing.

6 Lagged Interaction Features

- Combine GMP with other drivers to capture market dynamics:
 - $\text{GMP}_t * \text{Construction_Activity}_{t-1} \rightarrow$ economic demand vs. supply
 - $\text{GMP_growth}_{t-1} * \text{Employment_growth}_{t-1} \rightarrow$ captures productivity-driven absorption
 - $\text{GMP_per_capita} * \text{Population_growth} \rightarrow$ measures local purchasing power
- **Why:** Real estate markets are multi-dimensional — interactions improve predictive power significantly.

7 Economic Health & Overheating Indicators

- **Concept:** Detect if a metro is **overperforming or underperforming relative to trend**, which predicts future vacancy pressure.
- **Features:**
 - $\text{GMP_saturation} = (\text{GMP} - \text{GMP_trend}) / \text{GMP_trend}$
 - $\text{GMP_acceleration} \rightarrow$ change in growth rate
 - $\text{Volatility_index} = \text{std}(\text{GMP_growth_last_4Q}) \rightarrow$ economic stability measure
- **Why:** Shows overheating risk (high GMP above trend) \rightarrow potential rise in vacancy in near future due to oversupply.

8 Composite & Index Features

Top firms often build **single composite features** combining multiple signals:

- **Economic Health Score (EHS):**

$$EHS = 0.4 \times GMP_{Growth} + 0.3 \times (-GMP_{Volatility}) + 0.3 \times GMP_{relative_growth}$$

- **GMP Absorption Index:** Combines GMP growth, employment growth, and construction completion.
- **Why:** Reduces feature dimensionality while retaining predictive power.

9 Nonlinear Transformations

- **Log, Square Root, and Power Transforms** on GMP and growth rates to handle:
 - Wide differences in metro sizes
 - Nonlinear effects on vacancy
- **Top-notch example:**
 $Vacancy \sim \log(GMP)$ often captures diminishing marginal effect of economic output on vacancy.

10 Predictive Momentum & Turning Points

- **Top analysts detect inflection points in GMP trends:**
 - $GMP_growth_t * GMP_acceleration_t \rightarrow$ highlights shifts in momentum
 - Binary turning-point indicator $\rightarrow 1$ if GMP growth changes sign
- **Why:** Vacancy reacts to early signals before broader trends are apparent.

✓ Summary of Key “Top-Notch” Features

Feature	Type	Rationale
Lagged GMP (t-1, t-2)	Continuous	Captures delayed vacancy response
GMP_cycle_component	Continuous	Isolates cyclical deviations
GMP_relative_growth	Continuous	Measures metro vs national performance
Sectoral GMP ratios	Continuous	Links GMP to property type demand
GMP_elasticity	Continuous	Quantifies sensitivity of vacancy to GMP
Lagged interaction terms	Continuous	Captures multidimensional dynamics
GMP_saturation / Acceleration	Continuous	Detects overheating or emerging recovery
Composite indices (EHS, Absorption)	Continuous	Summarizes multiple economic signals
Nonlinear transformations	Continuous	Handles scale and diminishing returns
Turning-point indicators	Binary	Early warning signals for vacancy shifts

💡 **Insight:** Top real estate firms rarely use GMP in isolation. They **combine it with employment, construction, sectoral data, population growth, and lagged trends**, often producing **composite indices** that serve as leading indicators of vacancy.

If you want, I can now **draft a ready-to-use Python/Pandas implementation** for all these **top-notch features**, including sectoral ratios, lagged elasticities, cycles, and composite indices — fully aligned with industry best practices. This would be plug-and-play for your Moody's dataset.

Do you want me to do that?



So me some more features engineering with Gross_Metro_Product_Total.

Absolutely! Let's go even deeper — beyond standard growth rates, lags, and composite indices — into **cutting-edge, "out-of-the-box" feature engineering** for `Gross_Metro_Product_Total` (GMP) in the context of real estate vacancy modeling. These are the kinds of features **top real estate analytics teams** build.

◆ 1. Relative Trend Deviation Features

- **Concept:** Compare GMP to its long-term trend to detect overheating or underperformance.
- **How:**
 - $GMP_vs_Trend = GMP_t / \text{rolling_mean}(GMP, 8Q)$
 - $GMP_gap = GMP_t - \text{rolling_mean}(GMP, 8Q)$
- **Why:** Positive deviation → economy above trend → potential for rising vacancy if new supply overshoots demand.

◆ 2. Acceleration & Deceleration Features

- **Concept:** Capture changes in growth momentum.
- **How:**
 - $GMP_growth = (GMP_t - GMP_{t-1}) / GMP_{t-1}$
 - $GMP_acceleration = GMP_growth_t - GMP_growth_{t-1}$
- **Why:** Vacancy tends to react **after sustained acceleration/deceleration**.

◆ 3. Quarterly Seasonality Features

- **Concept:** GMP often has seasonal patterns (e.g., Q1 vs Q4 in certain metros).
- **How:**
 - $GMP_QoQ_Seasonal = GMP_t / GMP_{t-4}$
 - Encode **quarter of year** as categorical: Q1, Q2, Q3, Q4

- **Why:** Helps the model adjust for seasonal vacancy changes.

◆ 4. Rolling Volatility and Stability Features

- **Concept:** Measures how “stable” GMP growth is.
- **How:**
 - `GMP_volatility_4Q = rolling_std(GMP_growth, 4Q)`
 - `GMP_volatility_ratio = GMP_volatility_4Q / mean(GMP_growth, 4Q)`
- **Why:** High volatility markets often see unpredictable vacancy trends.

◆ 5. Cross-Market Relative Strength

- **Concept:** Rank metro GMP growth relative to other metros in the same quarter.
- **How:**
 - `GMP_rank = percentile_rank(GMP_growth_t across all metros)`
 - `GMP_rank_diff = GMP_growth_t - median(GMP_growth_all_metros)`
- **Why:** Relative strength signals **which metro's vacancy will outperform peers**.

◆ 6. GMP per Capita / Productivity

- **Concept:** Normalizes GMP by population or workforce size.
- **How:**
 - `GMP_per_capita = GMP / population`
 - `GMP_per_worker = GMP / total_employment`
- **Why:** Captures **efficiency and economic intensity**, which correlate with demand for space.

◆ 7. Economic Momentum Indicators

- **Concept:** Combine GMP growth and acceleration for a “momentum score.”
- **How:**
 - `GMP_momentum = GMP_growth * GMP_acceleration`
- **Why:** High momentum → metro may see **accelerating demand** and lower vacancy.

◆ 8. GMP Elasticity Proxies

- **Concept:** Measure **how sensitive vacancy is to GMP changes** historically.
- **How:**

- Rolling elasticity:

$$\text{Vacancy_elasticity_4Q} = \Delta \text{Vacancy} / \Delta \text{GMP over past 4 quarters}$$
- **Why:** Captures metro-specific responsiveness.

◆ 9. Lagged Cumulative Growth

- **Concept:** Measures cumulative economic growth over past N quarters.
- **How:**
 - $$\text{GMP_cum_growth_4Q} = (\text{GMP}_t / \text{GMP}_{t-4}) - 1$$
- **Why:** Vacancy reacts more strongly to **sustained growth** than single-quarter spikes.

◆ 10. Interaction with Other Features

- **Idea:** GMP alone is informative, but interactions add predictive power.
- **Examples:**
 - $\text{GMP_growth} \times \text{Construction_Activity} \rightarrow$ captures absorption vs. supply
 - $\text{GMP_per_capita} \times \text{Population_growth} \rightarrow$ potential housing demand
 - $\text{GMP_acceleration} \times \text{Employment_growth} \rightarrow$ indicates industrial/office space demand

◆ 11. Composite Indices

- **Economic Health Index:** Weighted combination of multiple GMP-based features:

$$EHI = 0.4 \times \text{GMP}_{growth} + 0.3 \times \text{GMP}_{acceleration} + 0.3 \times \text{GMP}_{volatility}^{-1}$$
- **Why:** Summarizes metro-level economic strength into one predictive metric for vacancy.

◆ Summary Table of Advanced GMP Features

Feature	Type	Purpose
GMP_vs_Trend	Continuous	Detect over/underperforming economy
GMP_acceleration	Continuous	Detect momentum shifts
GMP_QoQ_Seasonal	Continuous / Categorical	Adjust for seasonal effects
GMP_volatility_4Q	Continuous	Stability / uncertainty measure
GMP_rank	Continuous	Relative strength across metros
GMP_per_capita / GMP_per_worker	Continuous	Productivity / economic intensity
GMP_momentum	Continuous	Combined growth + acceleration
Vacancy_elasticity_4Q	Continuous	Metro-specific sensitivity
GMP_cum_growth_4Q	Continuous	Cumulative growth over time
GMP × Construction / Population	Interaction	Multi-dimensional demand signal

Feature	Type	Purpose
Economic_Health_Index	Composite	Aggregate predictive signal

These features **go far beyond simple growth rates** and are **used in institutional real estate modeling** to predict vacancy, absorption, and investment risk.

If you want, I can **create a Python implementation** for all these features ready to apply on your Moody's dataset, including **rolling windows, lag features, interactions, and composite indices**. This will be fully plug-and-play for your quarterly metro data.

Do you want me to do that next?