

# CPI-U 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 `consumer_price_index_urban` and response `vacancy`. Help me understand how it works

Perfect — let's unpack this carefully and connect **Consumer Price Index (CPI, specifically CPI-Urban)** with **vacancy rate** in the **real estate context**, especially since you're working with Moody's quarterly dataset across markets.

## 1. What is Consumer Price Index – Urban (CPI-U)?

**CPI-U** measures how prices for a fixed basket of goods and services change over time **for urban consumers** (which represent about 93% of the U.S. population).

It's the most common CPI metric reported by the Bureau of Labor Statistics (BLS).

In simple terms:

**CPI-U** ↑ → **Cost of living** ↑ → **Inflation** ↑

So, higher CPI-U means inflationary pressure — goods, services, and often construction costs become more expensive.

## 2. What is Vacancy Rate?

Vacancy rate = proportion of **rental or sale properties** that are **unoccupied** during a given period.

It tells us about **market tightness or softness**:

- **Low vacancy rate** → strong demand, low supply (tight market)
- **High vacancy rate** → weak demand, high supply (soft market)

## 3. Economic Link Between CPI-U and Vacancy Rate

Now, here's the interesting causal/economic relationship:

### A. Inflation → Rent and Cost Pressure

- Rising CPI often indicates **higher inflation**.
- Landlords may increase rents to **offset rising costs** (maintenance, utilities, property taxes, insurance).
- Higher rents can reduce **housing affordability**, especially in **lower-income urban areas**.

- This can lead to **higher vacancies**, as tenants move out or downsize.

**CPI↑ → Rents↑ → Affordability↓ → Vacancy↑**

## B. Inflation and Demand Side

- During high inflation, **interest rates** often rise (as central banks tighten policy).
- Higher interest rates make **mortgages more expensive**, reducing home purchases.
- That can **increase rental demand**, as people rent longer → **vacancy rate** ↓ in rental markets.

**CPI↑ → Interest rates↑ → Home buying↓ → Rental demand↑ → Vacancy↓**

So, **the effect is not one-directional** — it depends on which channel dominates:

Mechanism	CPI Impact on Vacancy	Typical in
Rent affordability channel	Vacancy ↑	High-rent or overbuilt markets
Substitution from ownership to renting	Vacancy ↓	Tight rental markets, rising mortgage rates



## 4. In Your Data Context (Quarterly Moody's Data)

Since your dataset has **multiple markets** and **quarterly frequency**, CPI-U changes can signal **local or national inflation waves** affecting **rents, construction, and affordability** differently across markets.

**Typical dynamics you might observe:**

Scenario	CPI_U trend	Vacancy trend	Interpretation
CPI_U rising fast, vacancy rising	Inflation is eroding affordability faster than demand can absorb supply	Rents too high, renters exit	
CPI_U rising, vacancy stable/falling	Inflation is pushing interest rates up, more renters stay in rental market	Rental market absorbing demand	
CPI_U falling, vacancy falling	Economy stabilizing, affordability improving	Healthy growth period	
CPI_U falling, vacancy rising	Economic slowdown or deflationary signal	Demand weakening overall	



## 5. Feature Engineering Ideas with consumer\_price\_index\_urban

To make this variable more powerful:

Feature	Description	Purpose
<b>cpi_yoy_change</b>	Year-over-year % change	Captures inflation acceleration
<b>cpi_qoq_change</b>	Quarter-over-quarter % change	Captures short-term inflation dynamics
<b>cpi_trend_4q</b>	Rolling 4-quarter average	Smooths noise, tracks trend
<b>cpi_volatility</b>	Rolling std dev over last 4 quarters	Measures inflation uncertainty
<b>real_rent_index</b>	Rent index adjusted for CPI	Measures <i>real</i> (inflation-adjusted) rent pressure
<b>cpi_market_diff</b>	Market CPI minus national CPI	Measures local inflation deviation

Feature	Description	Purpose
cpi_lag1, cpi_lag2	Lagged CPI	Captures delayed effect of inflation on vacancy

## 6. Practical Modeling Insight

When modeling **vacancy rate** with **CPI-U**:

- Expect **nonlinear or lagged effects** (CPI today affects vacancy next quarter).
- Consider **interaction terms** with:
  - **Interest rates** (CPI x InterestRate)
  - **Income growth** (CPI x PersonalIncome)
  - **Construction activity** (CPI x Housing\_Completions)

Because CPI's effect depends on **broader economic context** — inflation alone doesn't tell the full story.

Would you like me to show you **how to visualize this relationship** (e.g., CPI vs Vacancy with lag effects per market) or **how to design causal features** that could reveal inflation-driven vacancy shifts?



Help with out of box features engineering for this column.

Excellent — since you're working with **consumer\_price\_index\_urban (CPI-U)** in a **quarterly real estate dataset** (Moody's), let's go beyond the basics and design **out-of-the-box, industry-grade feature engineering ideas** that top real estate economists or data scientists might use to link inflation dynamics with **vacancy rates**.

We'll cover **economic, temporal, comparative, and behavioral** feature types.



## 1. Inflation Dynamics Features

These capture how fast and in what pattern inflation is changing.

Feature	Formula / Logic	Intuition
CPI_YoY_Change	% change vs same quarter last year	Measures overall inflation speed
CPI_QoQ_Change	% change vs previous quarter	Captures short-term shocks
CPI_Acceleration	Difference between current and previous QoQ change	Shows whether inflation is accelerating or decelerating
CPI_Trend_4Q	Rolling 4-quarter average of CPI	Smooths short-term volatility
CPI_Volatility_4Q	Rolling std. deviation of CPI (past 4 quarters)	High volatility indicates inflation uncertainty
CPI_Deviation_From_Median	CPI - median(CPI over 2 years)	Detects abnormal inflation episodes



**Use case:** When CPI volatility increases, investors may hesitate in development → new supply drops → vacancy can decrease later.

## 2. Lagged and Momentum Features

Inflation's effect on vacancy is rarely immediate — rents, construction, and household decisions adjust slowly.


Feature	Description
<b>CPI_Lag1, CPI_Lag2, CPI_Lag4</b>	CPI from 1, 2, and 4 quarters ago (to model delayed effects)
<b>CPI_Momentum_4Q</b>	Current CPI – CPI(4 quarters ago) → cumulative inflation over 1 year
<b>CPI_Reversal_Signal</b>	Sign change in inflation acceleration → indicates potential turning point

 **Tip:** You can use **lag features up to 1 year** to capture how prior inflation influences current vacancies.

## 3. Real-Adjusted and Derived Features

These translate CPI into *real-world affordability* or *real-term* measures.


Feature	Formula / Logic	Interpretation
<b>Real_Rent_Index = Rent_Index / CPI</b>	Adjusts nominal rent growth for inflation	Real purchasing power for renters
<b>Real_HPI = Home_Price_Index / CPI</b>	Real home price trend	Measures inflation-adjusted home affordability
<b>Real_Income = Personal_Income / CPI</b>	Inflation-adjusted income	If income growth < CPI → affordability worsens
<b>Affordability_Gap = Real_Income_Growth – CPI_YoY_Change</b>	Measures affordability pressure	
<b>Cost_Pressure_Index = (CPI × Construction_Cost_Index)</b>	Indicates broad inflationary stress on property markets	

 **Use case:** When CPI grows faster than real incomes, affordability declines → potential rise in **vacancy rates** for rentals.

## 4. Comparative Market Features

Capture how inflation in one market differs from the national or regional trend.

Feature	Formula / Logic	Use
<b>CPI_Local_vs_National = CPI_Market – CPI_National</b>	Market-specific inflation gap	Local inflation divergence
<b>CPI_Regional_ZScore</b>	$(\text{CPI\_market} - \text{mean\_region\_CPI}) / \text{std\_region\_CPI}$	Detects outlier inflation markets
<b>Inflation_Rank_Per_Quarter</b>	Rank of market's CPI vs others in that quarter	Relative inflation standing
<b>CPI_Gradient_Between_Markets</b>	CPI difference between neighboring markets	Cross-market pressure signal

 **Use case:** A market where inflation runs hotter than the national average may see reduced affordability faster → potential early warning for vacancies.

## 5. Interaction Features

These model *economic interplay* between inflation and other real estate or macro variables.


Feature	Formula / Logic	Insight
<b>CPI × Interest_Rate</b>	Inflation-monetary tension	Reflects cost of borrowing response
<b>CPI × Income_Growth</b>	Inflation-income alignment	Captures affordability resilience
<b>CPI × Housing_Completions</b>	Inflation-supply link	Rising costs may slow completions
<b>CPI × Employment_Rate</b>	Inflation-demand tension	Whether wage growth supports rent payments
<b>CPI / Unemployment_Rate</b>	Stagflation proxy	High CPI with low growth → high vacancy risk

 **Advanced idea:** Build a **nonlinear interaction** (like  $\log(\text{CPI}) * \text{InterestRate}$ ) for richer dynamics.

## 6. Behavioral / Expectation Features

These are more “creative” and forward-looking — often used by top real estate analytics firms.

Feature	Description	Why It's Useful
<b>Expected_Inflation_Proxy</b>	Use 1-quarter lead CPI (shifted backward) to simulate anticipated inflation	Markets react to <i>expected</i> , not current, inflation
<b>CPI_Shock_Dummy</b>	1 if CPI QoQ change > 90th percentile, else 0	Flags inflation shocks
<b>CPI_Regime_Label</b>	Cluster CPI trends into “Low / Moderate / High inflation regimes”	Model regime-dependent vacancy responses
<b>Inflation_Persistence_Index</b>	Ratio of current CPI trend to historical volatility	Measures inflation stickiness

 **Idea:** Inflation regimes could condition how sensitive vacancy is — e.g., in *high inflation* regimes, vacancy responds more to affordability than to supply.

## 7. Transformation Ideas

Sometimes, the best signal comes after reshaping the raw CPI time series.

Transformation	Effect
<b>Log(CPI)</b>	Stabilizes variance, better for regression
<b>First Difference (<math>\Delta\text{CPI}</math>)</b>	Focuses on inflation <i>changes</i>
<b>Second Difference (<math>\Delta^2\text{CPI}</math>)</b>	Captures acceleration/deceleration
<b>Z-normalization per market</b>	Removes cross-market level bias
<b>CPI_Seasonality_Adjusted</b>	Removes quarterly pattern, if any

## 8. Example Pipeline in Practice

If you were building this in Python:

python

```
df['cpi_yoy'] = df.groupby('market')['consumer_price_index_urban'].pct_change(4)
df['cpi_qoq'] = df.groupby('market')['consumer_price_index_urban'].pct_change(1)
df['cpi_acceleration'] = df['cpi_qoq'] - df.groupby('market')['cpi_qoq'].shift(1)
df['cpi_volatility_4q'] = df.groupby('market')['consumer_price_index_urban'].rolling(4).std().reset_index(0, drop=True)
df['cpi_lag4'] = df.groupby('market')['consumer_price_index_urban'].shift(4)
df['cpi_momentum'] = df['consumer_price_index_urban'] - df['cpi_lag4']
```

Would you like me to create a **shortlist of top 10 “high-impact” CPI features** (ranked by theoretical importance for vacancy modeling in real estate)?

That would help you prioritize which engineered features are most likely to improve your model.



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

Great — let's take a step up and identify **top-notch, industry-grade feature engineering ideas** specifically relevant for your scenario (quarterly real-estate dataset, multiple markets, response = vacancy rate) that leading real-estate analytics teams use. These will go beyond typical “lagged CPI” or “YoY change” features and delve into more advanced signals and modeling-ready constructs.

Here are **advanced feature engineering ideas** tailored to your context:

## 1. Market cycle / regime-features

- **Inflation-Regime Label:** classify quarters into regimes (e.g., *Low Inflation*, *Moderate*, *High/Accelerating*) based on your CPI variable. Then create a dummy or categorical feature indicating regime. The vacancy response may differ by regime.
- **Supply-Demand Imbalance Index:** for each market and quarter, compute something like:

$$\text{NewSupply} / \text{AbsorptionRate}$$

where NewSupply = completions or new units (if you have) and AbsorptionRate = previous quarter occupancy (or 1-vacancy). Combine with inflation: e.g., “Inflation × Supply-Demand Imbalance”.

- **Cycle Phase Proxy:** build something like (vacancy – long-term average vacancy) / std, or (rent growth – construction growth) as a phase-indicator. Then create feature: “CyclePhaseScore”. This helps reflect where the market is in its upward/downward phase.

## 2. Affordability & cost-pressure features (inflation focus)

- **Cost-Pressure Build Index:** Combine CPI growth + construction cost growth + labor cost growth (if available) → this signals rising cost base for property owners. Then derive feature of “CostPressureTrend\_QoQ” and “CostPressureTrend\_YoY”.
- **Affordability Gap Index:** (Income growth – CPI growth) as earlier + optionally (Rent growth – CPI growth). If income growth < CPI growth → affordability is worsening, which may push vacancy up

(or shift tenant composition).

- **Real Rent / Real Vacancy Ratio:** Realize rent growth adjusted for CPI ( $\text{rent\_index} / \text{CPI\_index}$ ). Then create features like “RealRentIndex\_YoY” and then compute “RealRentIndex\_YoY / Vacancy” as a derived ratio. This can capture how “real” rental cost is relative to vacancy.

### 3. Cross-market comparative features

- **Relative Inflation Spread:**  $\text{CPI\_market} - \text{CPI\_national}$  (or region). If a market’s inflation is significantly above national, that may stress local affordability.
- **Peer Market Momentum:** For each market, compute something like average of the other markets’  $\text{CPI\_YoY}$  or vacancy change, then create “Delta\_to\_Peer\_Avg =  $\text{CPI\_market\_YoY} - \text{PeerAvg\_CPI\_YoY}$ ”.
- **Market Divergence Score:** Combine several features: e.g.,  $(\text{Vacancy\_market} - \text{Vacancy\_national}) + (\text{CPI\_market} - \text{CPI\_national}) + (\text{IncomeGrowth\_market} - \text{IncomeGrowth\_national})$ . This gives a high-level signal of how “out-of-sync” a market is.

### 4. Temporal / lagged advanced features

- **Multi-horizon Lags:** Not only 1-quarter and 4-quarter lags, but also 2-quarter, 3-quarter, 8-quarter lags. Then features like “ $\text{CPI\_acceleration\_lag4} = (\text{CPI\_QoQ\_current} - \text{CPI\_QoQ\_lag4})$ ”.
- **Rolling Features:** Rolling window stats like 8-quarter rolling mean, standard deviation, skewness of CPI. For example: “ $\text{CPI\_roll\_std\_8q}$ ” or “ $\text{CPI\_roll\_skew\_8q}$ ”. High skew might indicate sudden inflation jumps → possible vacancy risk.
- **Change-of-Change Features:** Second derivative:  $\Delta^2\text{CPI} = (\text{CPI\_QoQ\_current} - \text{CPI\_QoQ\_lag1})$ . Then further interactions: “ $\Delta^2\text{CPI} \times \text{Vacancy\_lag1}$ ”.

### 5. Interaction features & non-linear transformations

- **CPI × Supply Growth:** If you have supply growth data (new construction completions), create “ $\text{CPI\_YoY} \times \text{SupplyGrowth\_YoY}$ ”.
- **CPI / Vacancy\_lag1:** To capture “given last quarter’s vacancy, how inflation this quarter might impact”.
- **Log/Exp Transforms:** Use  $\log(\text{CPI})$ ,  $\sqrt{\text{CPI}}$  if distribution is skewed; or use “ $\text{CPI\_ratio} = \text{CPI\_current} / \text{CPI\_lag4}$ ”.
- **Threshold/Dummy Features:** Create binary flags: “HighInflationShock” = 1 if  $\text{CPI\_QoQ} > 90\text{th percentile}$ ; “AffordabilityStress” = 1 if  $(\text{IncomeGrowth} - \text{CPI\_YoY}) < -x\%$ . These capture extreme events rather than smooth trends.

### 6. Derived target-leading features

- **Vacancy Leading Indicator:** Create a feature like “ $\text{CPI\_lag2\_to\_vacancy\_change} = \text{CPI\_lag2} / \text{Vacancy\_current}$  or something similar; this can embed the idea that inflation two quarters ago influenced current vacancy.
- **Market Entry / Exit Signal:** Derived from combinations: e.g., if  $(\text{CPI\_YoY} > 5\% \text{ AND } \text{SupplyGrowth\_YoY} > 3\% \text{ AND } \text{Vacancy\_lag1} < \text{MarketMedian})$  then mark as “RisingRisk” = 1. This is essentially a composite feature synthesising multiple signals.



## 7. Dimensionality reduction & latent features

- **Principal Components of Inflation + Macro + Market Variables:** Combine CPI, interest rates, income growth, employment growth, etc into a PCA and extract first few components (latent inflation/affordability pressure factor).
- **Clustering Features:** Cluster markets by inflation trajectory + vacancy trajectory + supply growth and then assign a “cluster label” as a feature (e.g., cluster 1 = high inflation & rising vacancy markets, cluster 2 = moderate inflation & low vacancy).
- **Graph/Spatial Features:** If markets have adjacency or similarity, build features like “average CPI of neighboring markets”, “spillover CPI” etc. (This is more involved but used in cutting-edge real-estate models). For example: in research you’ll see graph representation learning for real-estate appraisal.

[arXiv](#)

## 8. Alternative / unconventional data integrations

- **Sentiment or Consumer Confidence Feature:** Use consumer confidence index or expectation surveys; convert into “InflationExpectations” which may lead vacancy changes. According to one source, inflation influences cost of occupancy and supply cost. [plotzy.ai](#)
- **Remote Sensing / Night-lights Feature:** Especially for larger scale markets: use satellite night-light intensity changes in a region as a proxy for economic activity or occupancy stress. Research shows vacancy rate modelling using such data. [MDPI](#)
- **Digital Footprint / Mobility Feature:** Use mobility data (from public sources) as a proxy for tenant demand shifts (e.g., remote working reducing office occupancy → industrial/office vacancy dynamics).



### Prioritized List of “High-Impact” Features to Build First

Given your dataset and domain (CPI → Vacancy), here’s a recommended priority list:

1. `CPI_YoY_Change`, `CPI_QoQ_Change` (baseline)
2. `CPI_Lag2`, `CPI_Lag4`
3. `Affordability_Gap = IncomeGrowth_YoY - CPI_YoY_Change`
4. `Relative_Inflation = CPI_market - CPI_national`
5. `CPI_acceleration = CPI_QoQ_Change - CPI_QoQ_Change_lag1`
6. `CPI_roll_std_8q` and maybe `CPI_roll_skew_8q`
7. `HighInflationShock_Dummy` (`CPI_QoQ > percentile`)
8. `CPI × SupplyGrowth_YoY` (if supply data exists)
9. `InflationRegime_Label` (categorical)
10. `LatentInflationFactor` derived via PCA or clustering across markets

If you like, I can **prepare a Python snippet** (in pandas) with some of these advanced features already coded (for your quarterly multi-market dataset) which you can plug into your pipeline. Would that be



helpful?



So me some more features engineering with consumer\_price\_index\_urban.

Great — let's dive deeper and brainstorm **even more advanced / creative feature engineering ideas** centered on your variable Consumer Price Index – Urban (CPI-U) in your real-estate / vacancy modelling context. You can pick and choose the ones that make sense given your data availability and modelling goals.

## More Advanced Feature Ideas for CPI-U

Here are additional feature constructs you might build:

### 1. Inflation Expectation Gap

- $\text{CPI\_Forecast\_Error} = \text{CPI\_actual} - \text{CPI\_forecasted}$  (if you can derive or get a market forecast)
- $\text{CPI\_Expectations\_Dummy} = 1$  if  $\text{CPI\_YoY\_Change} > X\%$  AND previous quarter  $\text{CPI\_QoQ\_Change} < 0$  (indicating inflation reversal)
- Rationale: Markets often react to *unexpected* inflation rather than just inflation itself.

### 2. Real Cost of Holding / Ownership Pressure

- $\text{Real\_Cost\_Holding} = \text{CPI\_YoY\_Change} + \text{Property\_Tax\_Rate} + \text{Insurance\_Cost\_Growth}$
- $\text{Holding\_Cost\_Growth} = (\text{Construction\_Cost\_Index\_YoY} + \text{CPI\_YoY\_Change}) / 2$
- Link to vacancy: If holding cost rises sharply, owners might convert properties (vacancy impact) or delay maintenance.

### 3. Affordability Momentum Feature

- $\text{Affordability\_Momentum} = (\text{Income\_Growth\_YoY} - \text{CPI\_YoY\_Change}) - (\text{Income\_Growth\_YoY\_lag1} - \text{CPI\_YoY\_Change\_lag1})$
- This gives the **change in the gap** between income growth and inflation.
- If this momentum is negative (affordability deteriorating faster), vacancy might increase.

### 4. Real-Term Rent Pressure Feature

- $\text{Real\_Rent\_Growth} = \text{Nominal\_Rent\_Growth} - \text{CPI\_YoY\_Change}$
- $\text{Rent\_Inflation\_Ratio} = \text{Nominal\_Rent\_Growth} / \text{CPI\_YoY\_Change}$
- If real rent growth is negative, it may signal softness in rental demand → vacancy risk up.

### 5. Inflation Spike & Decay Features

- $\text{CPI\_Spike\_Dummy} = 1$  if  $\text{CPI\_QoQ\_Change} > \text{threshold}$
- $\text{CPI\_Decay\_Ratio} = (\text{CPI\_QoQ\_Change\_current} / \text{CPI\_QoQ\_Change\_lag1})$
- If a big spike and then a large decay, might signal correction phases in real-estate markets.

### 6. Market Saturation Pressure Index

- $\text{Inflation\_vs\_Supply} = \text{CPI\_YoY\_Change} - \text{SupplyGrowth\_YoY}$
- If inflation is high but supply growth is even higher (or similar), risk of oversupply (vacancy) increases.

## 7. Long-term Trend vs Short-term Shock

- $\text{CPI\_LongTrend} = \text{rolling\_mean}(\text{CPI\_YoY\_Change}, \text{window}=8 \text{ quarters})$
- $\text{CPI\_ShortTerm\_Shock} = \text{CPI\_QoQ\_Change} - \text{CPI\_LongTrend}/4$
- This helps separate structural inflation vs cyclical shock, which may have different impacts on vacancy.

## 8. Interaction with Monetary Environment

- $\text{Inflation\_Interest\_Spread} = \text{CPI\_YoY\_Change} - \text{Interest\_Rate\_change\_YoY}$
- $\text{Inflation\_Lease\_Clause\_Index} = \text{CPI\_YoY\_Change} * \% \text{ of leases with CPI escalation (if you have that)}$ . According to one analysis, leases often include CPI-linked clauses, so inflation → rental income effect. [investg.com](https://www.investg.com)<sup>1</sup>
- This interaction helps model how inflation couples with financing or lease structures that affect vacancy indirectly.

## 9. Spatial/Peer Inflation Pressure

- $\text{Peer\_CPI\_Avg} = \text{average CPI\_YoY\_Change of neighboring markets}$
- $\text{Inflation\_Spillover} = \text{CPI\_YoY\_Change\_market} - \text{Peer\_CPI\_Avg}$
- If your markets are Atlanta, Austin, New York etc., you can compute peer-inflation differential. A market with inflation much higher than peers may face local affordability or supply issues.

## 10. Regime-Switch / Cluster Features Based on Inflation

- Use clustering (k-means or hierarchical) on CPI trajectories (YoY, QoQ, vol) for each market to assign a “inflation trajectory cluster” label.
- Then use dummy features like `Cluster_HotInflation`, `Cluster_ModerateInflation`, `Cluster_LowInflation`.
- This approach is supported by literature where inflation regimes help explain real-estate dynamics. [cfe.columbia.edu](https://cfe.columbia.edu)<sup>1</sup>

## 11. Volatility & Uncertainty Features

- $\text{CPI\_Rolling\_StdDev\_4Q}$  (as you might already have)
- $\text{CPI\_Change\_Skew\_4Q}$  (skewness of quarterly CPI changes over last 4)
- $\text{CPI\_Uncertainty\_Index} = \text{Percentage of quarters in last year where } |\text{CPI\_QoQ\_Change}| > X\%$
- High inflation volatility or uncertainty may increase risk-aversion among tenants and owners, affecting vacancy.

## 12. Lead-Lag Relation Features

- $\text{CPI\_Lag2\_QoQ} = \text{CPI\_QoQ\_Change shifted by 2 quarters}$
- $\text{CPI\_Lag4\_YoY} = \text{CPI\_YoY\_Change shifted by 4 quarters}$
- Then create features like  $\text{Vacancy\_Pred\_Rel\_CPI\_Lag2} = \text{CPI\_Lag2\_QoQ} / \text{Vacancy\_current}$
- Helps capture delayed impact of inflation on vacancy.

## Practical Tips & Considerations

- **Scaling & normalization:** Because CPI levels differ by market, you might normalize features within each market (z-score) so that your model learns relative changes, not just absolute levels.
- **Avoiding multicollinearity:** Many engineered features will correlate (e.g., CPI\_YoY\_Change and CPI\_QoQ\_Change). Use feature-selection or regularization to avoid overfitting.
- **Time-series dependence:** Use proper train/test split (temporal split) and ensure lags don't leak future information.
- **Interpretability:** Some of these features (like clusters or latent factors) reduce interpretability. If you need client explainability, maintain a subset of "simple" features alongside complex ones.
- **Data availability:** Ensure you have the companion data to support interactions (e.g., interest rate changes, supply growth, lease escalation data). Without those, some features may be less reliable.
- **Lag structure analysis:** Empirically test which lags (lag1, lag2, lag4) of CPI features best correlate with vacancy movements in each market — the best lag may differ by market/property-type.

If you like, I can **generate a full list of ~20 feature engineering formulas (in pandas code)** using CPI\_U (and placeholders for companion variables) that you can plug directly into your pipeline. Would that be helpful?