

Consumer confidence and vacancy



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

Got it! Let's break this down carefully. You're working with **Moody's real estate dataset**, with quarterly observations per market, and your specific feature of interest is:

Feature: `consumer_confidence_index_urban`

Response: `vacancy` (likely the vacancy rate for properties in that market).

Here's a detailed explanation of how this feature might work and relate to vacancy rates:

1. What is Consumer Confidence Index (CCI) - Urban

- **Definition:** The Consumer Confidence Index measures the degree of optimism or pessimism that consumers feel about the overall economy and their personal financial situation.
- **Urban-specific:** Since this is `consumer_confidence_index_urban`, it specifically reflects **urban consumer sentiment**, which is more relevant to cities where commercial and residential real estate demand is concentrated.
- **Scale:** Typically, CCI is scaled so that 100 is neutral. Above 100 indicates optimism, below 100 indicates pessimism.

2. How CCI affects real estate vacancy

Think of vacancy rate as **the proportion of unoccupied units** in the market. CCI can influence this because:

1. Higher CCI → optimism → more spending → higher demand for property

- Consumers feel confident about income and jobs → more likely to rent/buy → vacancy decreases.

2. Lower CCI → pessimism → less spending → lower demand for property

- Consumers are cautious → fewer leases/sales → vacancy increases.

Expected relationship:

- In general, **negative correlation:** high consumer confidence → low vacancy; low confidence → high vacancy.
- But the magnitude may vary across markets, property types, and time periods.

3. Using this feature in a quarterly dataset

Since your dataset is **quarterly**:

- Consumer sentiment can **lead or lag vacancy**. People may take months to act on optimism/pessimism, so you could explore **lagged features**:
 - `consumer_confidence_index_urban_lag1` → CCI from the previous quarter
 - `consumer_confidence_index_urban_lag2` → CCI from 2 quarters ago
- This often improves predictive power because vacancy adjusts with delay to market sentiment.

4. Advanced feature engineering ideas (top real estate practice)

Here's where "out-of-the-box" comes in:

1. Growth Rate / Momentum:

- `cci_growth = (current_cci - previous_cci) / previous_cci` → captures acceleration in confidence.

2. Rolling averages:

- `cci_rolling3 = mean(cci of last 3 quarters)` → smooths short-term volatility.

3. Volatility / Uncertainty measure:

- `cci_std = std(cci of last 4 quarters)` → measures stability of consumer sentiment; volatile confidence may cause more conservative renting/buying → affects vacancy.

4. Interaction with other economic indicators:

- Example: `cci * unemployment_rate` → in cities with low unemployment, rising CCI may have stronger impact on vacancy.

5. Market segmentation:

- Separate models for different property types: offices, apartments, retail, industry. Urban CCI may impact apartments more than industrial properties.

5. Visualization/Exploratory Analysis

To understand the relationship:

- **Scatter plot:** CCI vs. vacancy
- **Lagged scatter:** Lagged CCI vs. vacancy
- **Correlation matrix:** Check correlation of CCI (and its transformations) with vacancy

✓ Summary:

`consumer_confidence_index_urban` is a forward-looking economic sentiment measure. Typically, **higher urban CCI → lower vacancy**, but lag effects and market specifics matter. Advanced features like growth rate, rolling averages, volatility, and interactions can improve predictive modeling.

If you want, I can **create a concrete feature engineering pipeline for `consumer_confidence_index_urban`** specifically tuned for quarterly real estate vacancy prediction. It would include lags, rolling averages, interactions, and volatility measures.

Do you want me to do that next?



Help with out of box features engineering for this column.

Perfect! Let's go beyond the standard lags and rolling averages and explore **"out-of-the-box" features** for `consumer_confidence_index_urban` (CCI) in real estate vacancy modeling. These are approaches top real estate data teams sometimes use to capture hidden signals:

1. Lagged Features with Flexible Windows

- Standard lags (1–2 quarters) are obvious. Out-of-the-box:
 - **Asymmetric lags:** e.g., `lag1`, `lag3`, `lag4` — testing different lag gaps because vacancy may respond non-linearly to confidence changes.
 - **Cumulative lag sum:** `cci_cum3 = CCI_lag1 + CCI_lag2 + CCI_lag3` → captures prolonged sentiment trends.

2. Rolling Dynamics

- **Rolling stats beyond mean:**
 - `cci_rolling_max_4 = max(CCI of last 4 quarters)` → captures recent peak optimism
 - `cci_rolling_min_4 = min(CCI of last 4 quarters)` → captures recent troughs
 - `cci_rolling_skew_4` → measures asymmetry of sentiment swings
 - `cci_rolling_kurt_4` → captures "spiky" confidence changes
- **Momentum features:**
 - `cci_momentum = CCI_current - CCI_lag4` → captures long-term acceleration in confidence

3. Rate of Change / Volatility

- **Quarter-to-quarter % change:** $(CCI_current - CCI_lag1) / CCI_lag1$
- **Standard deviation over rolling windows:**
 - `cci_volatility_4 = std(CCI over last 4 quarters)` → high volatility may signal uncertainty, affecting vacancy more than the absolute level.

4. Interaction Features

- **CCI × other economic indicators:** Vacancy is often influenced by combined economic conditions. For example:
 - **cci × urban_unemployment_rate** → optimistic consumers in low unemployment cities may drive stronger demand.
 - **cci × mortgage_rate** → rising CCI with low mortgage rates = stronger rental/buying activity.
 - **cci × office_rent_index** → interaction with property pricing dynamics.
 - **CCI × previous vacancy:** e.g., **cci_lag1 × vacancy_lag1** → captures market inertia; even with high CCI, high past vacancy may slow absorption.
-

5. Trend / Regime Features

- **Binary trend:**
 - **cci_trend_up = 1 if CCI_current > CCI_lag4 else 0** → identifies prolonged upward trends
 - **cci_trend_down = 1 if CCI_current < CCI_lag4 else 0**
 - **CCI state bins (quantiles):**
 - Convert CCI into **quartiles or deciles**: high, medium, low confidence zones
 - Helps non-linear models learn thresholds rather than assuming linear effect
-

6. Cyclical / Seasonal Features

- Consumer sentiment often has **seasonal effects**:
 - **cci_qtr_diff = CCI_current - CCI_same_quarter_last_year** → measures year-over-year changes
 - Useful for quarterly real estate datasets to remove seasonal bias
-

7. Advanced / “Exotic” Features

- **CCI trend acceleration:** **(CCI_current - CCI_lag1) - (CCI_lag1 - CCI_lag2)** → captures sudden change in sentiment trajectory
 - **CCI volatility clusters:** Flag sequences of high volatility → uncertainty periods may predict vacancy spikes
 - **Change of sign:** Count the number of times CCI crossed 100 in past year → signals market sentiment flip
-

✓ Summary

- Lags, rolling stats, momentum → capture **short- and medium-term effects**
- Interactions → capture **macro-economic dependencies**
- Trend / regime / volatility → capture **non-linear and market-specific effects**
- Seasonality / cyclical → capture **quarterly or yearly patterns**

These features often outperform simple lags in predictive power and explainability in real estate vacancy models.

If you want, I can **create a ready-to-use feature engineering table for consumer_confidence_index_urban** with 15–20 derived features specifically tuned for your quarterly real estate dataset. This would give you a concrete “production-ready” set to test in models.

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! Now I understand — you want **truly “top-notch” or industry-level feature engineering** for **consumer_confidence_index_urban** as used in real estate analytics. Let’s go beyond the usual lag/rolling features and explore what **leading real estate data science teams** actually do:

1. Leading/Lagging Indicators

- **Lead-lag analysis:** Vacancy doesn’t respond immediately to consumer confidence; top analysts create **lagged CCI features at multiple horizons**: 1, 2, 4 quarters.
- **Forward-looking expectations:** Some models use **CCI forecasts** (from economic reports) as features, not just historical CCI.
- **Cumulative effect:**

```
text

cci_cumulative_4q = sum(CCI of last 4 quarters)
```

Captures **sustained consumer optimism/pessimism**, which can more reliably predict vacancy changes.

2. Non-linear / Threshold Features

- **Binary or categorical bins:**
 - Divide CCI into quartiles or above/below 100. Helps capture **psychological thresholds** where consumer confidence meaningfully shifts behavior.
- **Change-of-state features:**
 - `cci_crossed_100 = 1` if CCI moved from `<100` to `>100`
 - Signals sentiment turning positive, often preceding drops in vacancy.
- **Saturation effects:**

- High CCI might not decrease vacancy further if the market is already tight — model non-linear effects via piecewise features.

3. Interaction Features with Other Economic Factors

- **Macro × micro interactions** are gold-standard in real estate modeling:
 - **CCI × Unemployment Rate** → optimism matters more when employment is stable
 - **CCI × Mortgage Rates** → high CCI + low rates → stronger leasing/buying activity
 - **CCI × Population Growth** → high confidence in growing urban areas = faster absorption
- **Property-type-specific interactions:**
 - Apartments vs Offices vs Retail: CCI might affect apartment demand faster than industrial demand.

4. Volatility & Trend Measures

- **Sentiment volatility:**
 - Standard deviation over last 4–8 quarters captures uncertainty. High volatility = cautious behavior → higher vacancy risk.
- **Trend acceleration:**

text

```
cci_acceleration = (CCI_current - CCI_lag1) - (CCI_lag1 - CCI_lag2)
```

Detects sudden spikes/drops in confidence that could precede market shifts.

- **Momentum / rate of change:** Quarter-over-quarter and year-over-year changes capture **speed of sentiment change**, not just level.

5. Seasonality & Cyclicity

- **Quarterly adjustment:** Real estate is seasonal; compare CCI to same quarter last year:

text

```
cci_qtrYoY = CCI_current - CCI_same_quarter_last_year
```

- **Rolling year features:** Average of last 4 quarters to smooth volatility and reveal underlying trend.

6. Regime-based Features

- **Market regimes:** Segment markets into regimes based on CCI trajectory:
 - Uptrend, Downtrend, Stable

- Combine with past vacancy to detect inertia:

```
text

cci_trend × vacancy_lag1
```

- **Confidence reversals:** Flag if CCI rapidly flips direction — top analysts often include this as a predictor of short-term volatility in occupancy.

7. Market Sentiment Derived Features

- **CCI relative to historical mean:**

```
text

cci_relative = CCI_current - mean(CCI over last 8 quarters)
```

Captures whether current sentiment is “unusually high or low” for that market.

- **CCI percentile:** Percentile rank within past N years — normalizes across markets with different baseline sentiment.

8. Advanced “Top-Tier” Approaches

- **Composite Indices:** Combine CCI with other leading indicators (retail sales, unemployment, wages) into a **market health index**. Often more predictive than CCI alone.
- **Dynamic Interactions:** Use rolling interactions:

```
text

rolling(CCI × Unemployment Rate, window=4q)
```

Captures changing macro-economic relationships over time.

- **Event-based adjustments:** Adjust CCI for known events (stimulus, policy changes) to capture their immediate impact on leasing demand.

✓ Summary

Top real estate teams rarely use raw CCI directly. Instead, they engineer **multi-horizon, non-linear, interaction-heavy, regime-sensitive features**, often combining CCI with other macro and market indicators. These features capture:

- Lagged effects
- Non-linear thresholds
- Volatility / acceleration
- Market interactions

- Seasonal/cyclical adjustments
- Composite economic signals

If you want, I can **prepare a full “production-ready” feature engineering blueprint** for `consumer_confidence_index_urban`, including **15–20 derived features** that incorporate all these top-tier strategies, ready for modeling vacancy.

Do you want me to do that next?



So me some more features engineering with `consumer_confidence_index_urban`.

Absolutely! Let's dive into **more advanced, “real-world” feature engineering ideas** for `consumer_confidence_index_urban` (CCI) specifically for quarterly real estate vacancy modeling. I'll focus on **creative, top-tier features** that go beyond simple lags and rolling averages.

1. Multi-horizon Lags & Differences

- `cci_lag1`, `cci_lag2`, `cci_lag4` → capture delayed response of vacancy to sentiment
- **Lag differences:**

text

```
cci_diff_1_4 = CCI_lag1 - CCI_lag4
```

Measures **medium-term sentiment change**, which can predict vacancy spikes.

2. Rolling Window Features

- **Rolling averages:** `cci_rolling_mean_4q = mean(CCI last 4 quarters)`
- **Rolling max/min:**

text

```
cci_rolling_max_4q = max(CCI last 4 quarters)
cci_rolling_min_4q = min(CCI last 4 quarters)
```

Captures recent **peaks and troughs in optimism**.

- **Rolling volatility / std:**

text

```
cci_rolling_std_4q = std(CCI last 4 quarters)
```

High volatility often signals **uncertainty** → **higher vacancy**.

3. Momentum / Acceleration

- **Quarter-to-quarter % change:**

```
text

cci_pct_change_1q = (CCI_current - CCI_lag1) / CCI_lag1
```

- **Acceleration:** Detect sudden changes in sentiment trajectory:

```
text

cci_acceleration = (CCI_current - CCI_lag1) - (CCI_lag1 - CCI_lag2)
```

4. Trend & Regime Indicators

- **Binary trend flags:**

```
text

cci_uptrend = 1 if CCI_current > CCI_lag4 else 0
cci_downtrend = 1 if CCI_current < CCI_lag4 else 0
```

- **Threshold-based features:**

```
text

cci_above_100 = 1 if CCI_current > 100 else 0
cci_below_100 = 1 if CCI_current < 100 else 0
```

- **Change-of-state:** Detect shifts from negative → positive sentiment.

5. Interaction Features

- **CCI × other macro factors:**
 - CCI × unemployment_rate → optimism matters more when employment is high
 - CCI × mortgage_rate → high confidence + low rates = faster leasing
- **CCI × past vacancy:**

```
text

cci_lag1 × vacancy_lag1
```

Captures market inertia; even high confidence may not immediately reduce vacancy if market is already saturated.

6. Seasonal / Cyclical Features

- **YoY comparison:**

```
text

cci_yoy = CCI_current - CCI_same_quarter_last_year
```

Removes seasonal bias and highlights growth/decline in sentiment.

- **Rolling yearly average:**

```
text

cci_rolling_mean_4q = mean(CCI over last 4 quarters)
```

Smooths short-term volatility and highlights trend.

7. Relative / Normalized Features

- **CCI relative to historical mean:**

```
text

cci_relative = CCI_current - mean(CCI_last_8q)
```

Signals whether current sentiment is unusually high or low for that market.

- **CCI percentile rank:**
 - Converts raw CCI into **market-relative percentile**, useful when comparing multiple markets.

8. "Exotic" / Industry-Level Features

- **Volatility clusters:** Count sequences of high volatility → predicts instability in leasing demand
- **Confidence reversals:** Flag quarters where CCI sharply changes direction
- **Composite indices:** Combine CCI with other leading indicators (wages, unemployment, retail sales) → more predictive than single indicator
- **Market-adjusted CCI:**

```
text

cci_adjusted = CCI_current - mean(CCI_market_history)
```

Adjusts for baseline differences between markets (NY vs Austin vs Atlanta).

✓ Key Insight:

Top real estate analysts rarely use **CCI in isolation**. Instead, they extract:

1. **Temporal signals** (lags, rolling, momentum, acceleration)
2. **Threshold/non-linear signals** (trend flags, above/below 100, reversals)
3. **Market interactions** (with vacancy, unemployment, rates)
4. **Seasonal/cyclic signals** (YoY change, rolling mean)
5. **Normalized signals** (percentiles, market-adjusted)

These features allow vacancy models to detect **both immediate and delayed effects of consumer sentiment**, even in noisy quarterly data.

If you want, I can **create a full “feature blueprint” table for `consumer_confidence_index_urban`**, listing 20+ features with exact formulas and explanations, ready to plug into your model.

Do you want me to do that next?