DANA 4830 – PROJECT

Objective 1: One-Week-Ahead Occupancy Prediction Using Seasonal-Trend Decomposition

Author – Sunaina Jain

Introduction

This report presents a one-week-ahead occupancy prediction using Seasonal and Trend decomposition (STL) on a synthetic building dataset provided by the US Department of Energy. Inspired by Alaraj et al. (2023), the main objective is to replicate their work but for a shorter-term forecast—one week ahead instead of one year. Our goal is to assess and analyze occupant behavior to improve HVAC electricity consumption.

To explore the effect of seasonal variation, we considered **two separate timeframes**:

- **January–March 2005 (Spring)**
- **October–December 2005 (Fall)**

Methodology

We followed the exact STL-based forecasting steps outlined in the referenced research paper. The workflow is:

1. **Filtered for working days only** to ensure predictions are based on actual operational days.
2. **Extracted occupancy data** separately from Jan–Mar and Oct–Dec 2005.
3. Applied **STL decomposition** on the 10-minute interval time series using a period of 144 (one day = 24 hours × 6 intervals).
4. Isolated the **trend and seasonal components** of occupancy behavior.
5. **Forecasted occupancy** for January 1–7, 2006 by repeating the seasonal pattern and adding the average trend.
6. **Visualized and compared** the predicted occupancy to the actual values from January 1–7, 2005 to evaluate performance.

This prediction provides a foundational understanding of occupancy patterns, which directly affect HVAC system usage and thus future energy demand.

1. Occupancy-based Prediction Framework

1.1 Load and Preprocess Data

- Extract 3 months of occupancy data (working days only).
- Drop weekends and holidays using SiteDayTypeIndex.
- Calculate **total occupants** by summing across rooms.
- Create a datetime column with 10-minute intervals.

1.2 STL Decomposition

- Apply STL with **seasonal period = 24×6 = 144** (1-day seasonality at 10-min resolution).
- Extract trend and seasonal components.
- Drop residuals.
- Average trend and reuse seasonality.

1.3 One-week-ahead Prediction (Jan 1–7, 2006)

- Reconstruct predicted values:
  occupancy_pred = avg_trend + seasonal [:1008]
  (1008 intervals = 7 days × 24 hours × 6).

---

2. Comparative Analysis

2.1 Baseline: Actual Jan 1–7, 2005

- Extract the same week from 2005.
- Sum across rooms → total_occupants.
- Use as **baseline for comparison**.

2.2 Compare 3 Predictions

- Jan–Mar 2005 based prediction
- Oct–Dec 2005 based prediction
- Actual Jan 1–7, 2005
  Line plots for each with consistent time axis (normalized).

Data Source

- File Path: s3://oedi-data-lake/building_synthetic_dataset/A_Synthetic_Building_Operation_Dataset.h5
- Variable: ZonePeopleOccupantCount/block1_values
- Frequency: 10-minute intervals

About Data

```
✅ File Opened Successfully!
📁 Available Group: 1. README
📁 Available Group: 2. Resources
📁 Available Group: 3. Data
🔍 Exploring '1. README' ...
    🔢 Dataset '1. README' found!
🔍 Exploring '2. Resources' ...
    📁 Group '2. Resources' contains: ['2.1. Building Models', '2.2. Weather Files']
🔍 Exploring '3. Data' ...
    📁 Group '3. Data' contains: ['3.1. Metadata', '3.2. Timeseries']
```

```
✅ File Opened Successfully!

📂 Datasets in 'Building Models':
- MediumOfficeDetailed_90.1-2013_1A_High_efficiency.osm
- MediumOfficeDetailed_90.1-2013_1A_Low_efficiency.osm
- MediumOfficeDetailed_90.1-2013_1A_Standard_efficiency.osm
- MediumOfficeDetailed_90.1-2013_3C_High_efficiency.osm
- MediumOfficeDetailed_90.1-2013_3C_Low_efficiency.osm
- MediumOfficeDetailed_90.1-2013_3C_Standard_efficiency.osm
- MediumOfficeDetailed_90.1-2013_5A_High_efficiency.osm
- MediumOfficeDetailed_90.1-2013_5A_Low_efficiency.osm
- MediumOfficeDetailed_90.1-2013_5A_Standard_efficiency.osm

📂 Datasets in 'Weather Files':
- 1A
- 3C
- 5A

📂 Datasets in 'Metadata':
- brick_relationhsips.json
- brick_relationhsips.ttl

📂 Datasets in 'Timeseries':
- 1A
- 3C
- 5A
```

```
📂 Listing Datasets Inside: 3. Data/3.2. Timeseries/1A
📌 High
📌 Low
📌 Standard
```

Listing Datasets Inside: 3. Data/3.2. Timeseries/1A/Standard
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
TMY3

Listing Datasets Inside: 3. Data/3.2. Timeseries/1A/Standard/2005
run_1
run_2
run_3
run_4
run_5

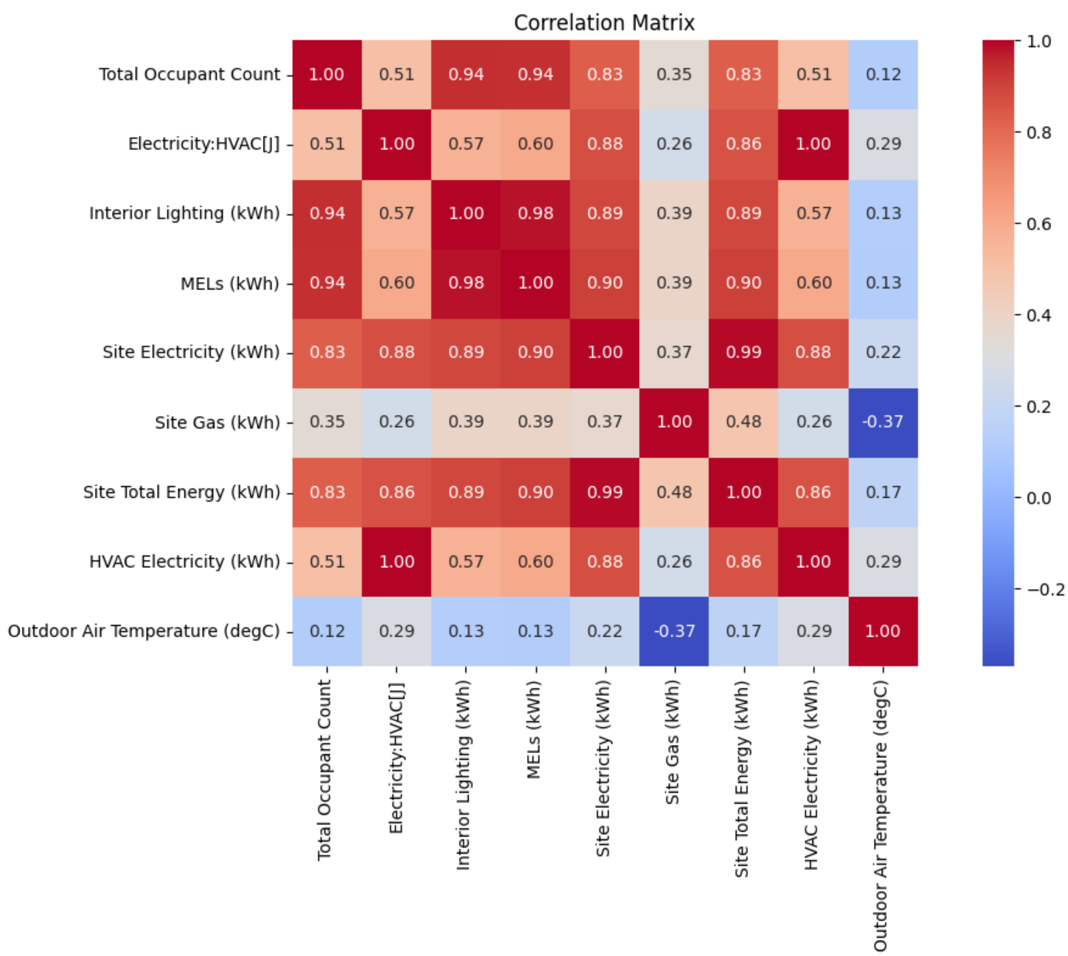
```
Listing Datasets Inside: 3. Data/3.2. Timeseries/1A/Standard/2005/run_1
AirSystemOutdoorAirEconomizerStatus
CoolingElectricity
ElectricityFacility
ElectricityHVAC
ExteriorLightsElectricity
FanAirMassFlowRate
FanElectricPower
FansElectricity
GasFacility
GasHVAC
HeatingElectricity
InteriorEquipmentElectricity
InteriorLightsElectricity
PumpElectricPower
PumpMassFlowRate
PumpsElectricity
SiteDayTypeIndex
SiteHorizontalInfraredRadiationRateperArea
SiteOutdoorAirDewpointTemperature
SiteOutdoorAirDrybulbTemperature
SiteOutdoorAirRelativeHumidity
SiteOutdoorAirWetbulbTemperature
SystemNodeMassFlowRate
SystemNodePressure
SystemNodeRelativeHumidity
SystemNodeTemperature
ZoneAirRelativeHumidity
ZoneAirTerminalVAVDamperPosition
ZoneElectricEquipmentElectricPower
ZoneLightsElectricPower
ZoneMeanAirTemperature
ZoneMechanicalVentilationMassFlowRate
ZonePeopleOccupantCount
ZoneThermostatCoolingSetpointTemperature
ZoneThermostatHeatingSetpointTemperature
```

```
Listing Datasets Inside: 3. Data/3.2. Timeseries/1A/Standard/2005/run_1/CoolingElectricity
axis0
axis1
block0_items
block0_values
block1_items
block1_values
```

Description of Variables:

AirSystemOutdoorAirEconomizerStatus	Shows if the outdoor air economizer is active (0/1)
CoolingElectricity	Electricity used by cooling systems (Joules)
ElectricityFacility	Total electricity consumption of the facility (Joules)
ElectricityHVAC	Electricity consumed by HVAC systems (Joules)
ExteriorLightsElectricity	Electricity used for exterior lighting (Joules)
FanAirMassFlowRate	Air mass flow rate through fans (kg/s)
FanElectricPower	Power consumed by fans (Watts)
FansElectricity	Total electricity used by fans (Joules)
GasFacility	Total gas consumption in the building (Joules)
GasHVAC	Gas consumption by HVAC systems (Joules)
HeatingElectricity	Electricity used for heating (Joules)
InteriorEquipmentElectricity	Electricity used by internal equipment (Joules)
InteriorLightsElectricity	Electricity used by interior lighting (Joules)
PumpElectricPower	Power used by pumps (Watts)
PumpMassFlowRate	Mass flow rate of pump systems (kg/s)
PumpsElectricity	Electricity consumption of pumps (Joules)
SiteDayTypeIndex	Index indicating the type of day (weekday/weekend)
SiteHorizontalInfraredRadiationRateperArea	Infrared radiation per unit area (W/m²)
SiteOutdoorAirDewpointTemperature	Dewpoint temperature outside (°C)
SiteOutdoorAirDrybulbTemperature	Outdoor air temperature (°C)
SiteOutdoorAirRelativeHumidity	Relative humidity outside (%)
SiteOutdoorAirWetbulbTemperature	Wet bulb temperature (°C)
SystemNodeMassFlowRate	Mass flow rate in HVAC nodes (kg/s)
SystemNodePressure	Pressure in HVAC system nodes (Pa)
SystemNodeRelativeHumidity	Relative humidity at HVAC system nodes (%)
SystemNodeTemperature	Temperature at HVAC system nodes (°C)
ZoneAirRelativeHumidity	Relative humidity inside building zones (%)
ZoneAirTerminalVAVDamperPosition	Position of VAV dampers (0-1 scale)
ZoneElectricEquipmentElectricPower	Power used by electric equipment in zones (Watts)
ZoneLightsElectricPower	Power used by lights in zones (Watts)
ZoneMeanAirTemperature	Mean air temperature inside zones (°C)
ZoneMechanicalVentilationMassFlowRate	Mass flow rate of mechanical ventilation (kg/s)
ZonePeopleOccupantCount	Number of occupants in each zone
ZoneThermostatCoolingSetpointTemperature	Thermostat cooling setpoint (°C)
ZoneThermostatHeatingSetpointTemperature	Thermostat heating setpoint (°C)

Basic Explanatory Data Analysis



- Total Occupant Count has a very strong correlation with:
  - Interior Lighting (0.94)
  - MELs (Misc. Electric Loads) (0.94)

This makes sense — more occupants mean more lights on and devices in use.

- Site Electricity vs. Total Energy (0.99):

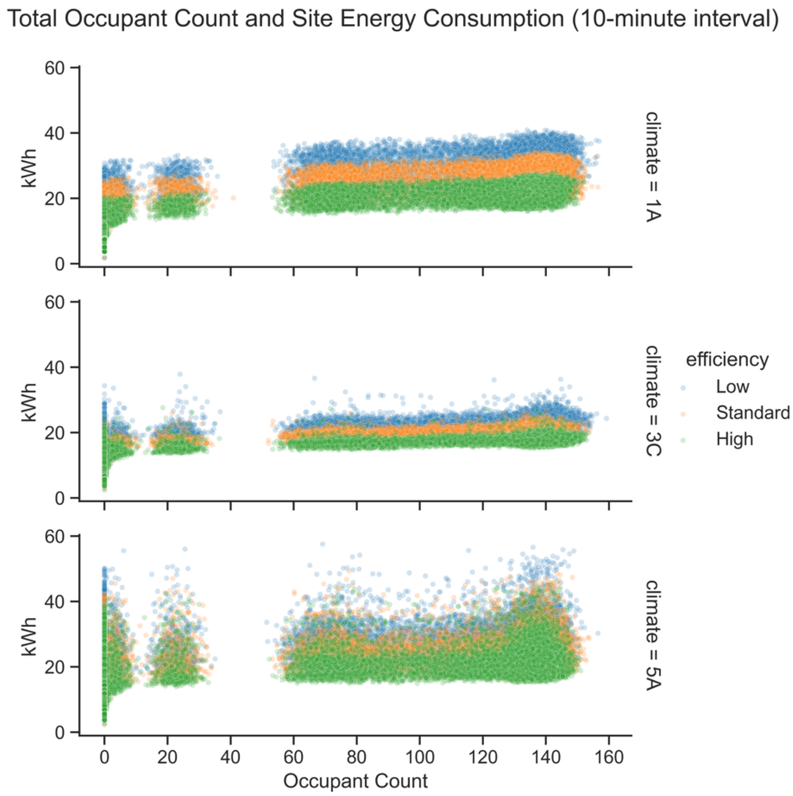
Almost all Site Total Energy is coming from electricity, not gas.

- Electricity: HVAC is strongly tied to:
  - Site Electricity (0.88)
  - Site Total Energy (0.86)
  - Occupant Count (0.51) → moderately strong

Shows HVAC usage significantly contributes to overall energy use and scales with people in the building.

- Negative correlation with Site Gas (-0.37):

Possibly indicates gas is used for heating, so as outside temperature rises, gas use drops.



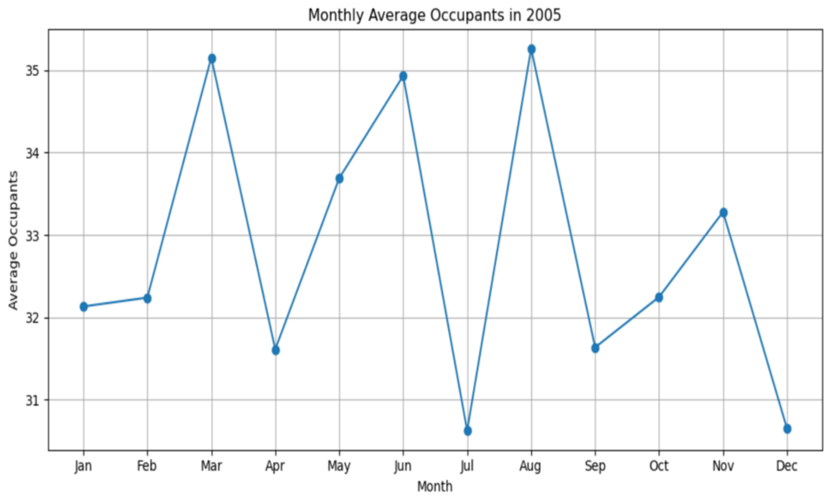
Across all three climates, there is a clear positive relationship between occupant count and site energy usage — as more people occupy the building, total energy use increases. The red regression lines (Low, Standard, High) show different slopes and intercepts, indicating that energy efficiency levels impact how energy scales with occupancy:

- High Efficiency (green) buildings consume less energy for the same number of occupants.
- Low Efficiency (blue) buildings use more energy even at similar occupant levels.

- (1A) Miami (hot-humid): Higher energy use overall, especially when occupancy increases.
Strong upward slope — energy demand is very sensitive to occupancy here.
Efficiency separation is clear: High efficiency saves a lot of energy even at high occupancy.

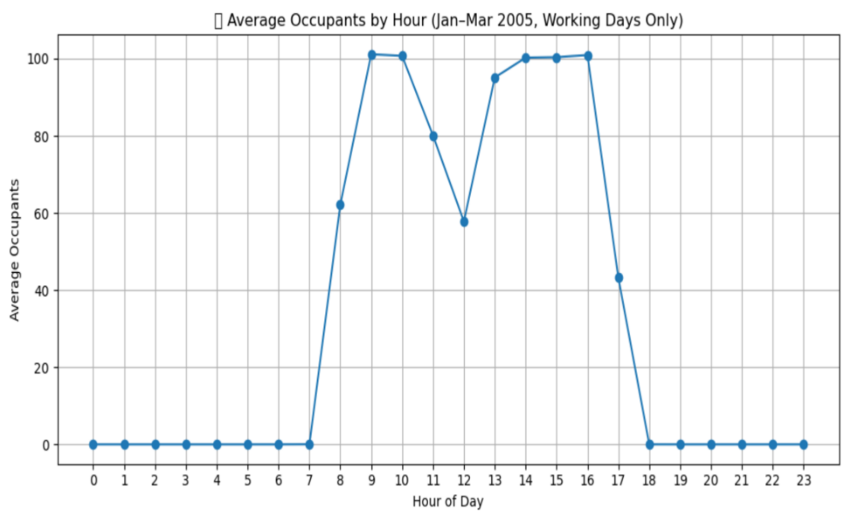
- (3C) San Francisco (mild): Lower overall energy use.
Slope is still positive but gentler.
Efficiency levels still make a difference, though the spread between them is slightly narrower than in 1A.

- (5A) Chicago (cold): Energy use increases with occupancy, but the pattern looks a bit more dispersed.
Possibly more variation due to heating loads dominating at low occupant counts.
Efficiency benefits still hold: High-efficiency buildings consistently use less energy.

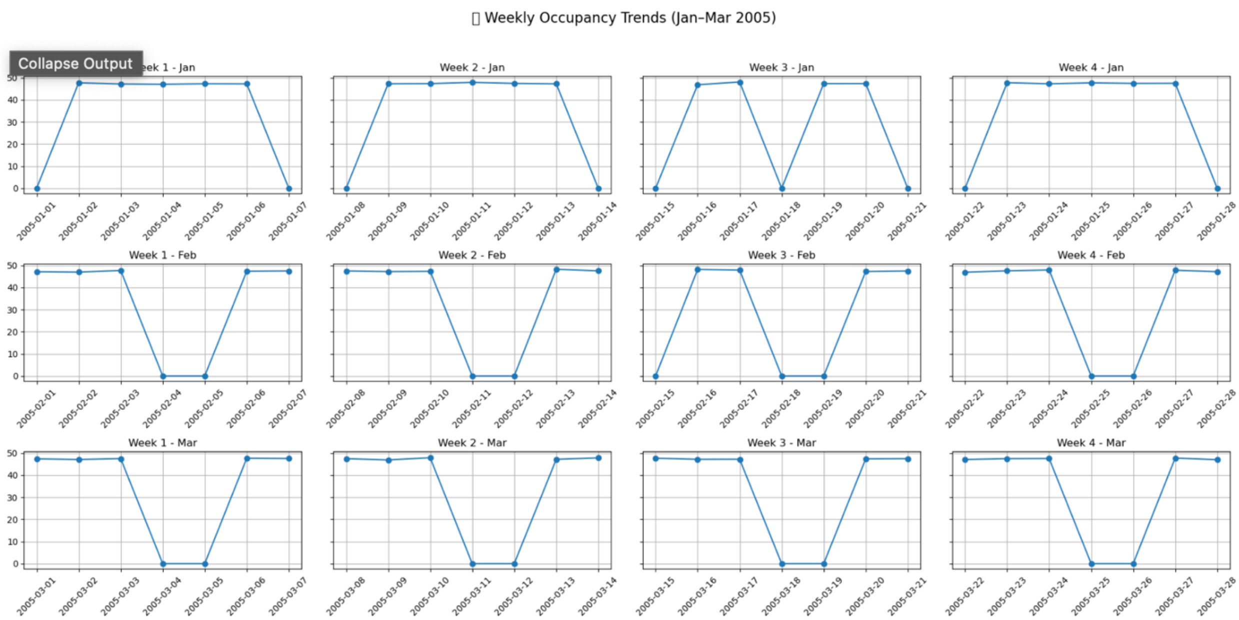


This consistency suggests that occupant behavior does not heavily depend on seasonal variations, validating the assumption made in the original research paper that the nature of building usage remains consistent across the year.

The slight dips in April, July, and December may be attributed to potential holidays or breaks, while peaks in March and August could reflect periods of increased office utilization. Nevertheless, the overall pattern reinforces the idea that weekday operations dominate occupancy dynamics, making one-week-ahead forecasting feasible without major seasonal adjustments.



The figure illustrates the average number of occupants in the building by hour of the day during working days from January to March 2005. As seen in the trend, occupancy begins to rise sharply after 7:00 AM, reaching peak levels around 9:00 AM. This peak is maintained until 12:00 PM, with a slight drop during lunch hours, and rises again until 5:00 PM. After 6:00 PM, occupancy rapidly drops to zero, confirming that building usage follows a standard work schedule (8 AM to 6 PM). These occupancy dynamics are essential to accurately estimating HVAC energy consumption, as heating and cooling loads are directly influenced by the number of people in the building during operational hours.



Across all weeks, a clear working-day occupancy pattern is visible:

- **Monday to Friday** show high occupancy (≈ 45–50 occupants on average).
- **Weekends (Saturday & Sunday)** drop sharply to near-zero occupancy.

This confirms the building was used only on weekdays, supporting earlier findings.

This strong repetitive structure supports STL decomposition and justifies forecasting future occupancy by **repeating seasonal weekday trends** and averaging the trend component.

Hence, we continued it by removing weekends.

Step 1: Load Jan–Mar 2005 Data & Apply STL (seasonal period = 144)

STL is a method that breaks down a time series into three parts:

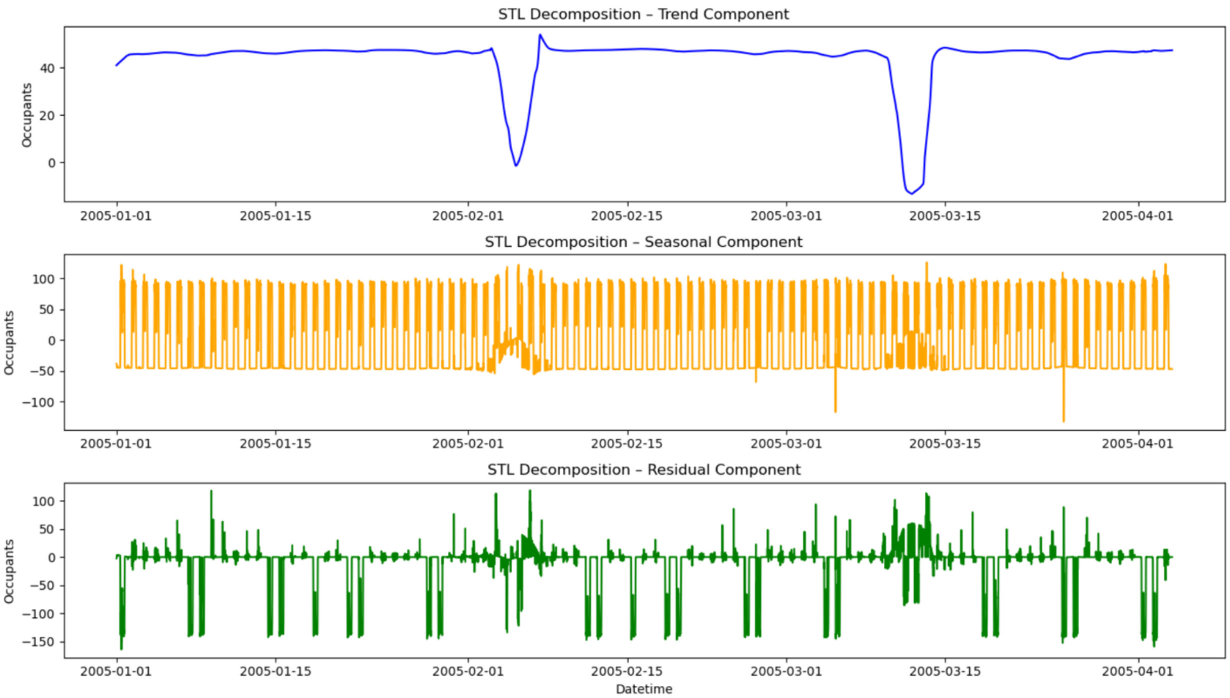
- **Trend** – long-term progression (e.g., building gets busier over months).
- **Seasonal** – repeating patterns (e.g., daily work hours, weekly cycles).
- **Residual** – leftover noise or random variation not explained by the above.

We are working with 10-minute interval data, and we want to capture daily cycles in occupancy.

- There are 6 intervals per hour (60 min / 10 min = 6).
- A day has 24 hours × 6 = 144 intervals.

So setting seasonal=144 tells STL to look for patterns that repeat every 144 time steps**,** i.e., once per day**.**

Step 2: Visualize STL Components – Trend, Seasonal, Residual

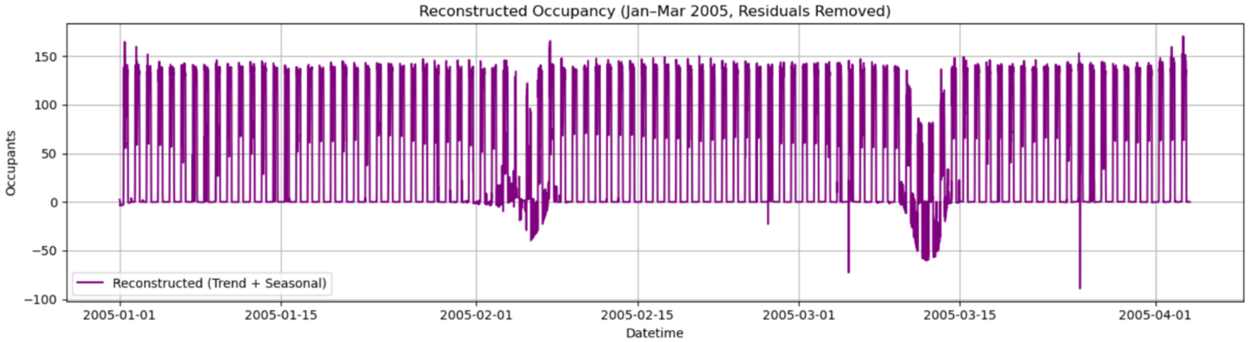


- Repeating daily occupancy cycles, extracted by STL with a 1-day (144-point) period.
- The second chart shows the **consistent daily behavior** across multiple days. Regular peaks and troughs every 144 data points (1 day = 24 hours × 6 intervals of 10 min). Peaks during office hours (e.g., 9 AM – 5 PM), flat at night and weekends. It is **very predictable and reusable**. This is the core of occupancy behavior we'll reuse for forecasting. It reflects employee schedules, work routines, and is seasonally consistent.

- The third chart is remaining noise after subtracting trend and seasonality. Captures unusual **events**, measurement errors, and anomalies.
- Random, irregular spi1kes.
- Higher activity around holidays or weekends.
- **Very noisy and unpredictable**. However, we discard this for prediction, since it's random and not meaningful for long-term forecasting.

Overall, this decomposition gives us **control over patterns** in our time series. By keeping only **trend + seasonal** components, we build a **clean and realistic forecast**. This approach ensures our **HVAC occupancy prediction** is smooth, repeatable, and energy efficient.

Step 3: Drop Residuals and Retain Only Trend + Seasonal Components



This line plot shows the **reconstructed occupancy time series** for the first 3 months of 2005 — after performing **STL decomposition** and dropping residuals.

- **Y-axis**: Number of total occupants across all zones.
- **X-axis**: Date and time from Jan 1 to Mar 31, 2005 (at 10-minute intervals).
- **Line (purple)**: Sum of Trend + Seasonal components (residuals excluded).
- The plot shows clear daily occupancy cycles, with sharp rises and drops — indicating a regular working schedule (e.g., 9 a.m. to 6 p.m.).
- There are small dips in February and March, possibly due to mid-week holidays or unusual patterns (still captured by trend).
- Random noise and outliers are removed, so the signal looks cleaner and smoother than the original data.

This reconstructed signal is smoother**,** removes holiday/sick-day noise. It reflects the true working-day occupancy behavior. This will be used in the next steps to generate predictions for Jan 1–7, 2006**.**

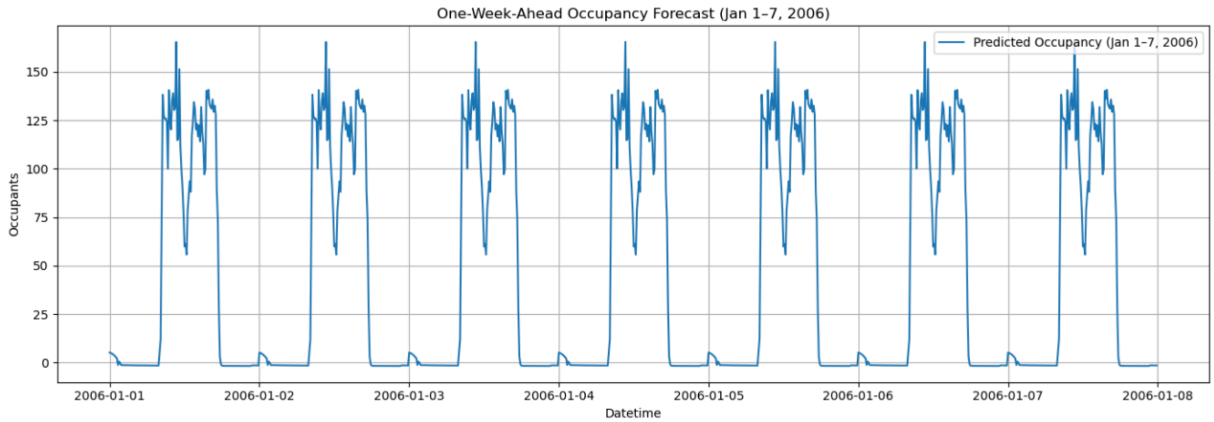
Step 4: Average the Trend and Predict Occupancy

Use the cleaned STL components to forecast occupancy for a future week (e.g., Jan 1–7, 2006). We assume:

- The **trend** is **stable** and can be approximated by its **average**.
- The **seasonal pattern** (daily cycles) is repetitive and can be **reused**.
- Residuals are dropped (already done in Step 3).

The trend component reflects long-term direction (e.g., increase or decrease over months). Since we are predicting just one week ahead, we can safely assume a stable baseline. So instead of extrapolating trend using a model (which may overfit or mislead), we average it to get a robust constant level for the next week.

```
Average trend value: 43.80371762942844
```



This chart visualizes the **forecasted number of occupants for each 10-minute interval** during the first week of **January 2006**, using:

- The **average trend** value (≈ 43.8 occupants)
- A repeated **daily seasonal pattern** extracted from Jan–Mar 2005
- **Residuals removed** (assumed to be noise)

Consistent Daily Patterns

- Each day follows nearly the same shape and range.
- High occupancy during working hours (approx. 8 AM – 6 PM).
- Near-zero during nighttime and early morning hours.

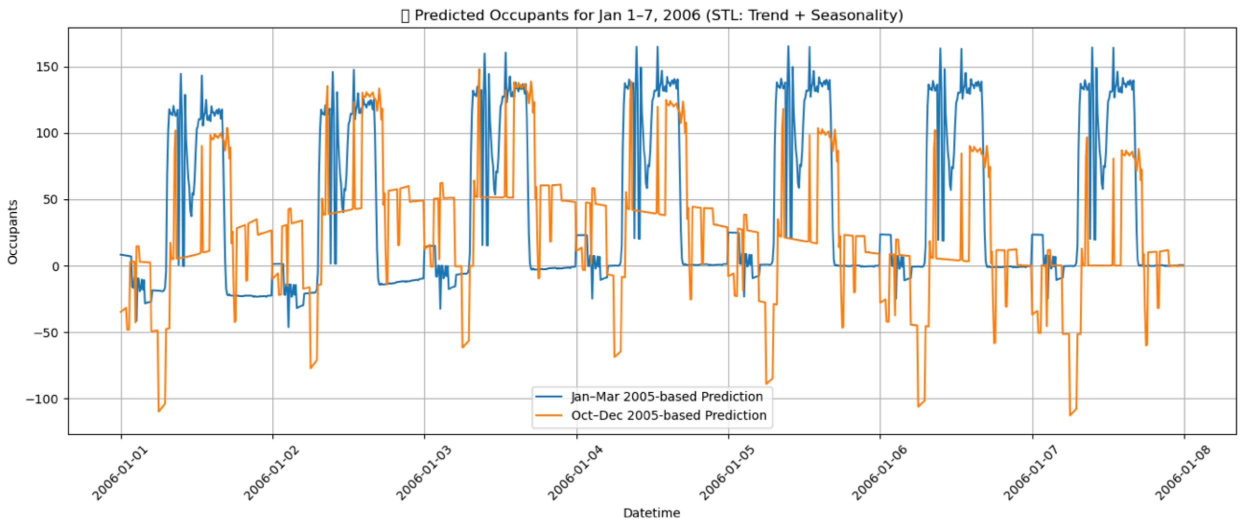
Clear Peaks & Valleys

- Peaks exceed **160 occupants** — likely office full capacity.
- Sudden drops correspond to off-hours (nights/weekends).

Repetition Across Days

- Identical shape for each day reflects our assumption:
  Future behavior = past daily cycle + stable average trend

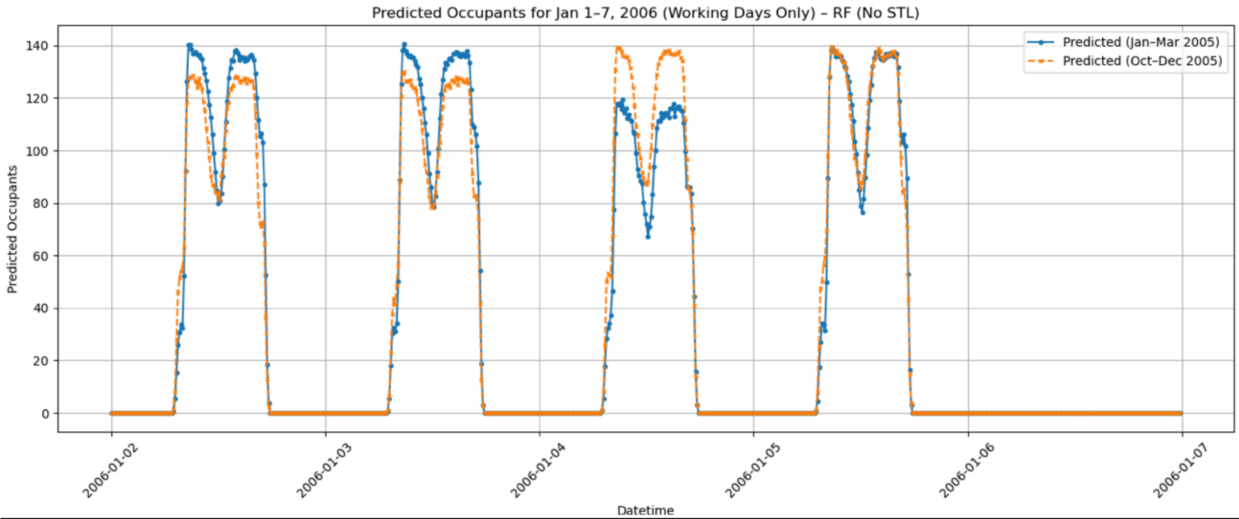
Prediction of JAN2006 from Fall as well as spring



Two prediction lines:

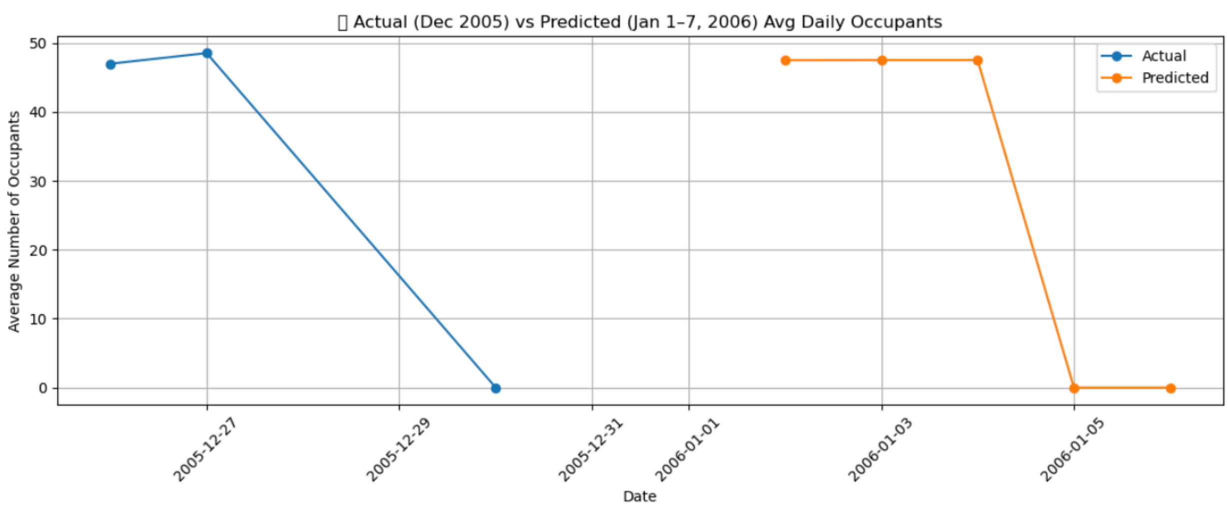
1. **Blue Line – Jan–Mar 2005-based Prediction:**
   - This uses STL decomposition (trend + daily seasonality) from **Winter (Jan–Mar)**.
   - The pattern here is more **stable** and represents expected winter weekday work cycles.
2. **Orange Line – Oct–Dec 2005-based Prediction:**
   - Uses STL decomposition from **Fall (Oct–Dec)**.
   - This line shows **more noise**, possibly due to holidays, breaks, or seasonal transition effects in the training data.

Without stl



Feature	Jan–Mar Model	Oct–Dec Model
Peak Occupancy	Slightly higher	Slightly lower
Early Week (Jan 2–3)	More optimistic	Slightly conservative
Later Week (Jan 4–6)	Both align closely	Both align closely
Start/End of Day	Sharp drop-offs (0)	Identical behavior

Random Forest (No STL) yields more robust and accurate predictions, especially with less noise. STL helps understand time-series components but depends heavily on the quality of past data. In our case, STL with Jan–Mar 2005 performs decently, but with Oct–Dec 2005, the prediction suffers.



There's a dip during the New Year transition, followed by a partial rebound. The predicted trend mirrors pre-holiday occupancy, implying the model expected a return to normalcy post holidays.

This highlights how historical holiday behavior and calendar effects play a big role in occupancy modeling.

Interestingly, predictions using **STL decomposition from Jan–Mar (spring)** data yielded more reliable results than those from **Oct–Dec (fall)**, possibly due to fewer holidays and a more stable working schedule. This confirms that **seasonal consistency** in training data enhances forecast accuracy. In contrast, the **Random Forest model** provided more flexible and often more accurate predictions without requiring time-series decomposition but lacked interpretability.

In conclusion, integrating occupancy forecasting into building management systems can **reduce HVAC electricity use**, lower carbon footprints, and improve comfort levels. Even a **one-week-ahead forecast**—as shown in this project—can empower facility managers to **pre-emptively adjust HVAC setpoints**, turning occupant data into actionable energy intelligence.