Analyzing Traffic Occupancy Rate Patterns on San Francisco Freeways

Divya Varshini Suravaram • Veera Vivek Telagani • Soumith Reddy Podduturi

Topics for the day

1. **Dataset Overview**

   Summary of the data structure, sensors, and time intervals

2. **Data Visualization**

   Visual exploration of daily and hourly traffic patterns.

3. **Model Training & Evaluation**

   Training models to predict traffic occupancy and assess performance

4. **Trend Analysis**

   Seasonal decomposition and weekly occupancy trends

Project Focus and Objectives

This project focuses on analyzing occupancy data from sensors to identify usage patterns, peak times, and trends, enabling data-driven insights for optimizing space management.

Dataset Overview

Dataset Features

963

Sensors

From different car lanes on freeways in the San Francisco Bay Area

455

Days of data

From January 1, 2008 to March 30, 2009

144

Time intervals per day

10-min time intervals in a day

2

Anomalies

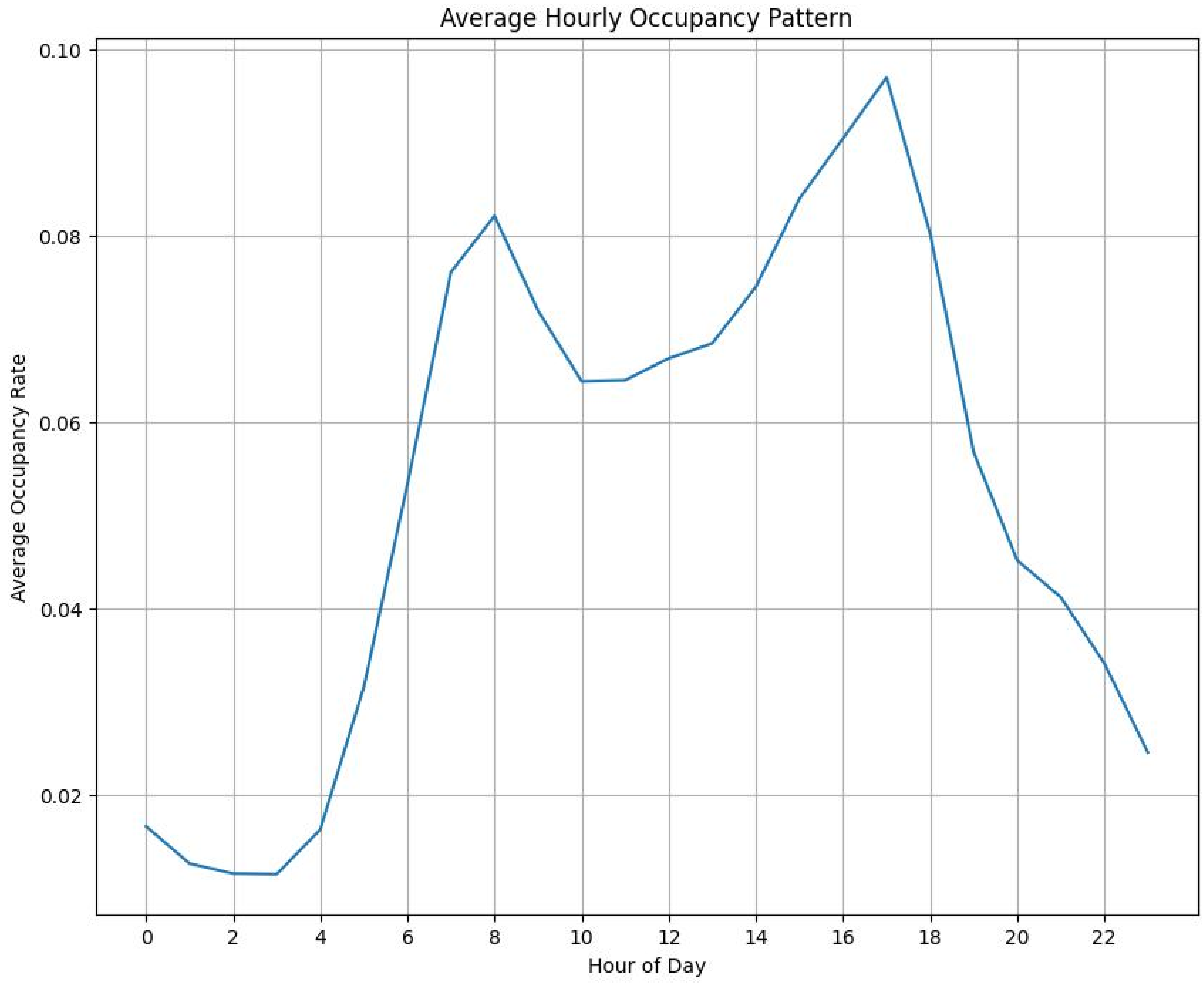
All sensors were muted between 2:00 and 3:00 AM on March 8, 2008 and March 9, 2009

Data Visualization

Data Visualization

Average Hourly Occupancy Pattern

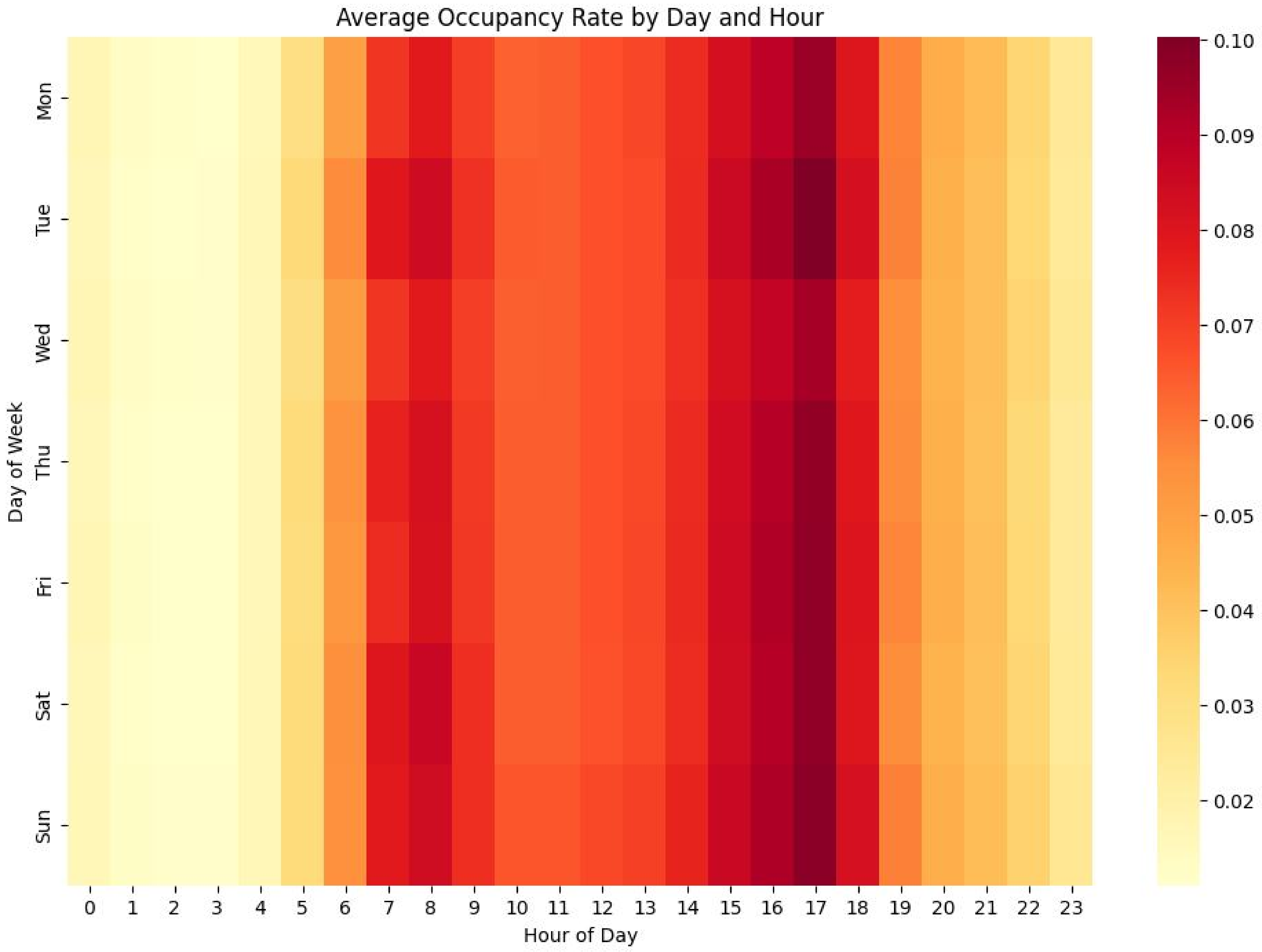
The graph highlights peak traffic occupancy around 8 AM and between 4 PM to 6 PM, reflecting typical office hours. These patterns indicate increased road usage during morning and evening commutes, primarily driven by office-goers.



Data Visualization

Average Occupancy Rate by Day and Hour

Higher occupancy rates are observed during weekday mornings around 7 AM to 9 AM and evenings between 4 PM to 6 PM, indicating typical commute hours. Weekends show relatively lower occupancy, reflecting reduced work-related traffic.

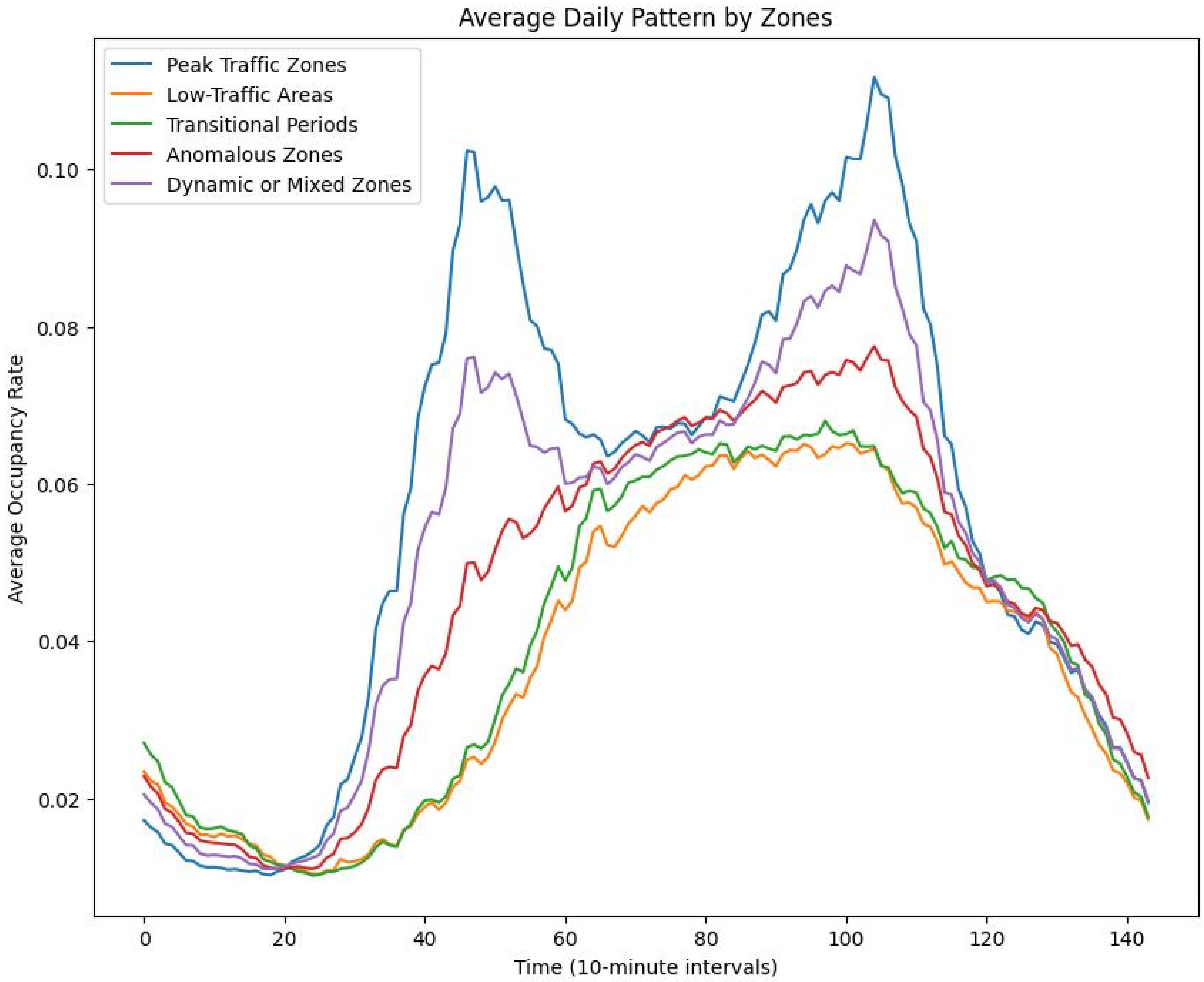


Data Visualization

Average Daily Pattern by Zones

Each cluster represents distinct traffic behaviors:

- **Peak Traffic Zones** have consistently high occupancy, particularly during morning and evening commute hours.
- **Low-Traffic Areas** show minimal activity throughout the day.
- **Transitional Periods** exhibit moderate and steady traffic.
- **Anomalous Zones** display irregular or unique patterns.
- **Dynamic or Mixed Zones** show fluctuating occupancy, indicating variable usage.



Model Training & Evaluation

Feature Extraction

Morning Rush

Average occupancy during morning rush hours.

Evening Rush

Average occupancy during evening rush hours.

Mid-day Rush

Average occupancy during midday hours.

Night Rush

Average occupancy during night hours.

Hourly Means

Mean traffic occupancy for each hour.

Hourly Variation

Hourly variation in traffic occupancy.

Rush Ratio

Ratio of morning to evening occupancy.

Day-Night Ratio

Ratio of day to night traffic.

First Hour

Occupancy during the first hour of data.

Last Hour

Occupancy during the last hour of data.

Peak Variation

Difference between max and min occupancy.

Steps for training a Model

1. Feature Scaling

The **StandardScaler()** in the pipeline scales the features, ensuring all variables have a mean of 0 and a standard deviation of 1. This helps models like Gradient Boosting and Random Forest perform optimally.

2. Training the Model

The pipeline fits both **Random Forest** and **Gradient Boosting** models on the training data (X_train and train_labels) as part of the Voting Classifier ensemble.

3. Making Predictions

The trained pipeline predicts labels for the test data (X_test) using the ensemble's soft-voting mechanism, combining the probabilities from Random Forest and Gradient Boosting.

Model Parameters

RandomForestClassifier

```
n_estimators=2000,
max_depth=None,
min_samples_split=2,
min_samples_leaf=1,
max_features='sqrt',
bootstrap=True,
class_weight='balanced_subsample',
random_state=42,
n_jobs=-1
```

GradientBoostingClassifier

```
n_estimators=1000,
learning_rate=0.005,
max_depth=10,
min_samples_split=4,
min_samples_leaf=2,
subsample=0.8,
random_state=42
```

VotingClassifier

```
estimators=[
    ('rf', randomForestClassifier),
    ('gb', gradientBoostingClassifier)
],
voting='soft',
weights=[2, 1]
```

Evaluation Metrics

Evaluation metrics show **92.49%** accuracy with strong performance across all days of the week

0.92
Accuracy

0.93
Precision

0.92
Recall

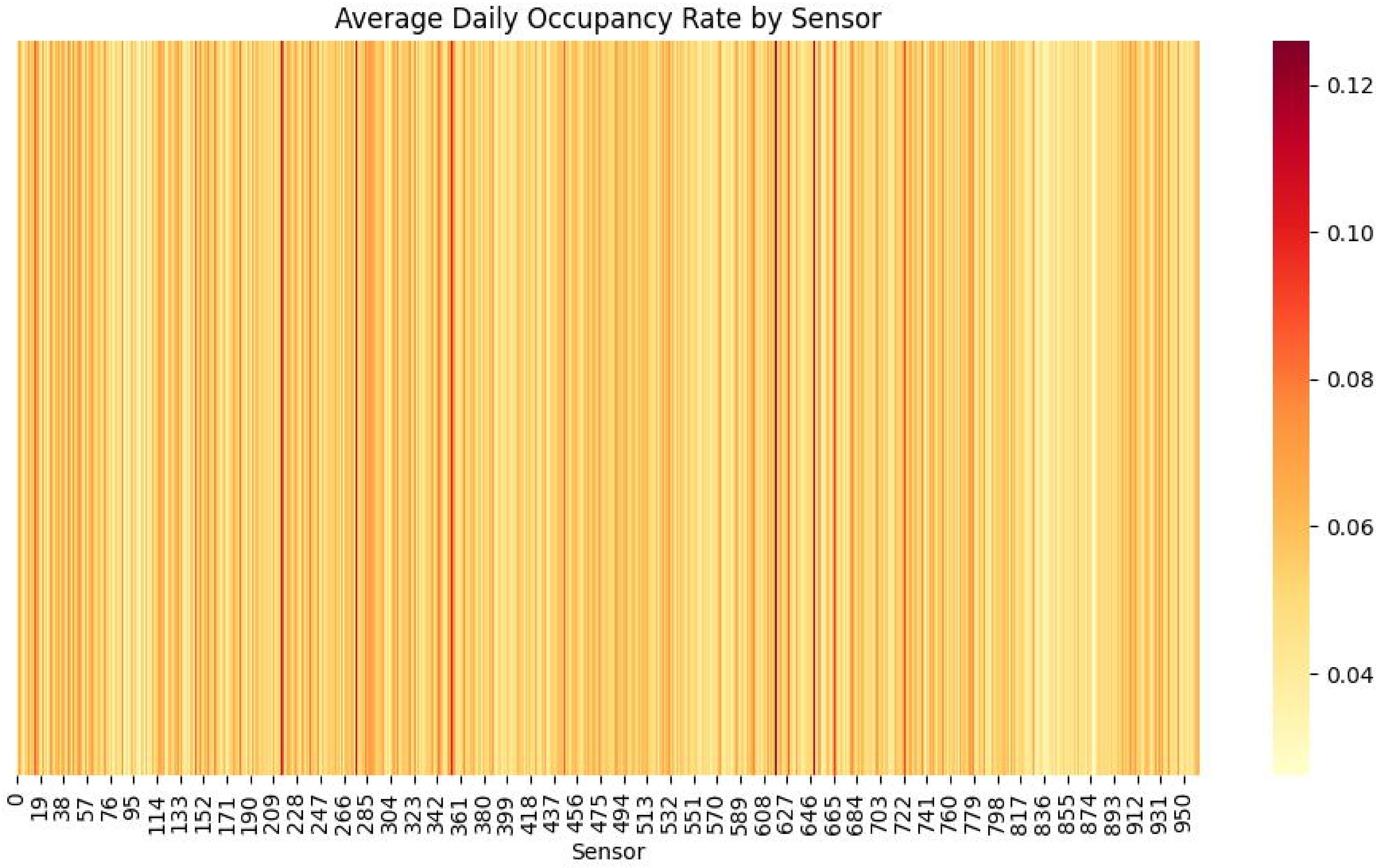
0.92
F1 Score

Trend Analysis

Trend Analysis

Average Daily Occupancy Rate by Sensor

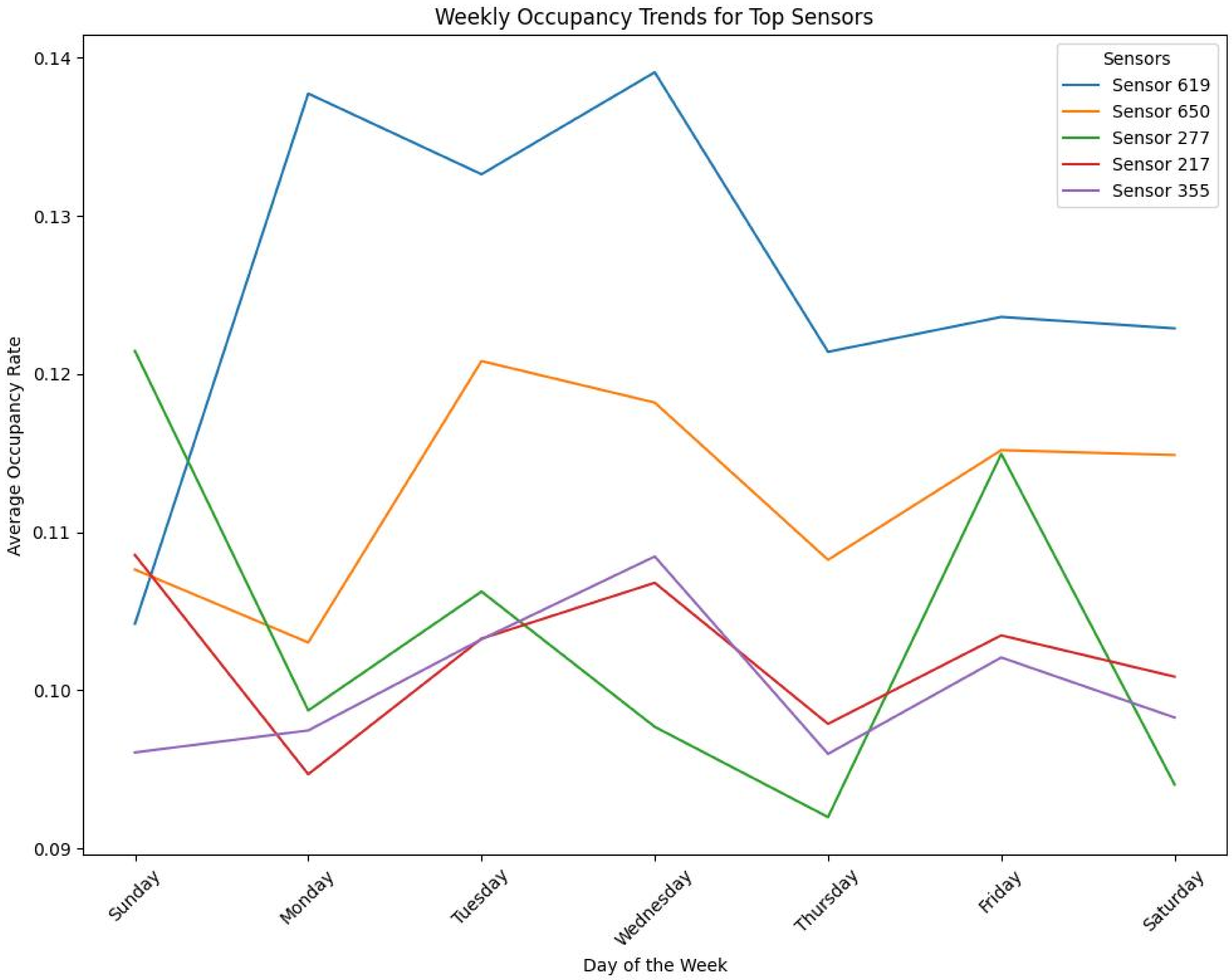
The heatmap illustrates the average daily occupancy rates for individual sensors. Darker shades indicate higher occupancy for those particular sensors, while lighter shades represent sensors with lower occupancy rates.



Trend Analysis

Weekly Occupancy Trends for Top Sensors

The line chart displays the weekly occupancy trends for the top sensors. Each line represents a specific sensor, with fluctuations showing how traffic occupancy changes across the days of the week. Peaks and dips indicate higher and lower occupancy rates for these sensors, highlighting variations in traffic patterns by day.

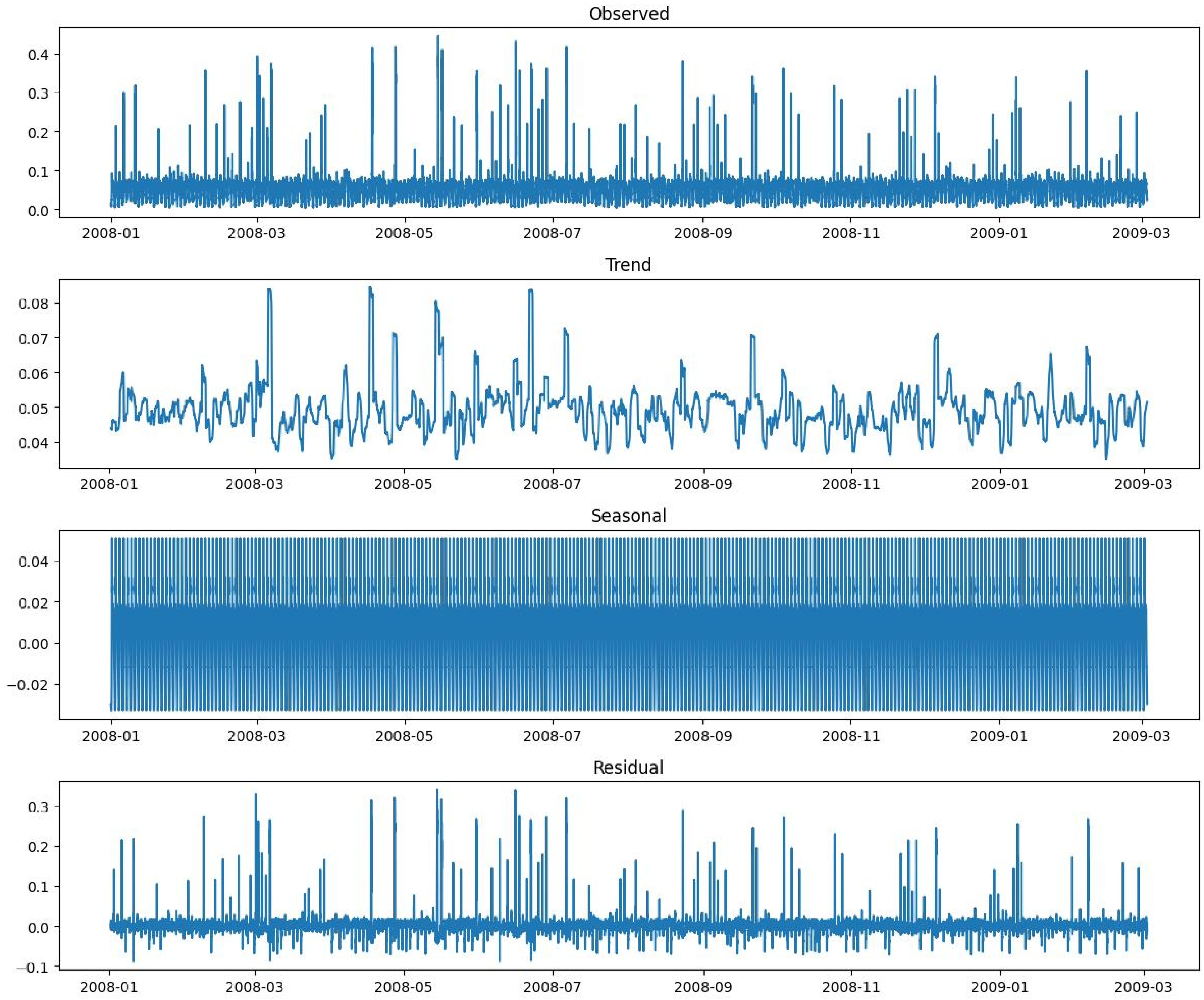


Trend Analysis

Seasonal Decomposition

This visualization shows the seasonal decomposition of traffic occupancy data

1. **Observed**: Original data with all patterns combined.
2. **Trend**: Long-term traffic patterns over months.
3. **Seasonal**: Repeating daily or weekly traffic fluctuations.
4. **Residual**: Irregular variations or anomalies after removing trends and seasonality.

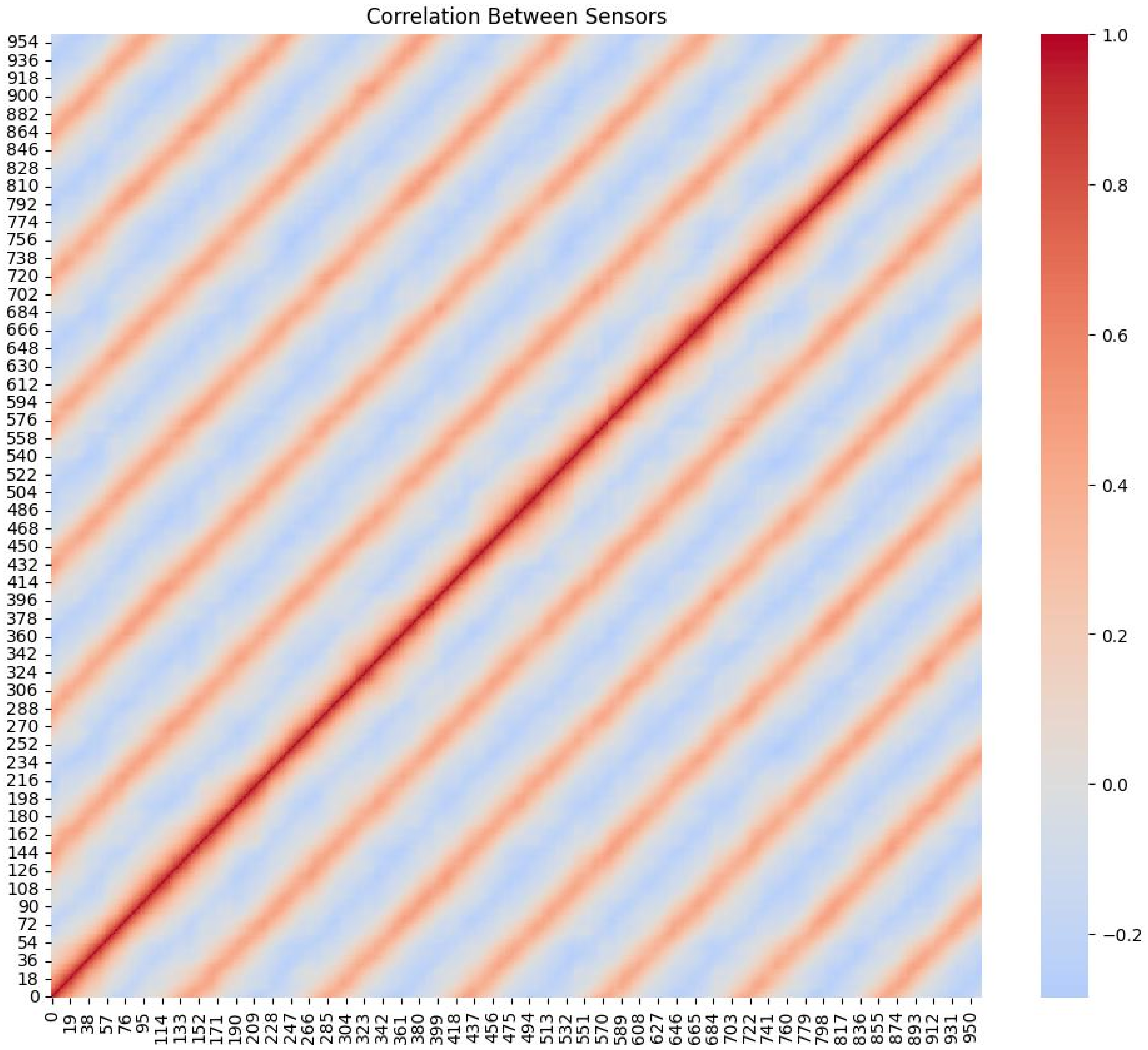


Trend Analysis

Correlation Between Sensors

This heatmap shows the correlation between sensors based on their traffic occupancy data.

- **Red areas (positive correlation)**: Sensors with similar traffic patterns, likely monitoring nearby or related locations.
- **Blue areas (negative correlation)**: Sensors with opposing traffic trends, possibly in different zones or time frames.
- The diagonal line represents each sensor's correlation with itself, which is always 1.



Conclusion

This project successfully analyzed traffic occupancy data to uncover meaningful patterns, including peak hours, daily trends, and anomalies. The methods and visualizations validated the approach, demonstrating clear insights that can support better traffic management and planning decisions.

Thank you