



澳門大學
UNIVERSIDADE DE MACAU
UNIVERSITY OF MACAU

Usage Analysis and Recommendations for Sports Facilities at the University of Macau

Group Members: Liang Zhongkai He Jingping Zhang Yi Yao Jiahao Liang Zhengping

2025.12.13

PART 1

1 Motivation

2 Data Collection

3 Data Preprocessing and Analysis

4 Total court analysis

PART 2

5 Badminton Court Data Analysis Process

6 Tennis Court Data Analysis Process

7 Squash Court Data Analysis Process

8 Machine learning



Background & Pain Point

Sports venues are core scenarios for students' health management, post-class relaxation and social interaction, but there is a problem of difficult booking during evening and busy academic periods.

Core Objective

Identify peak crowded periods and easily accessible times of venues, and explore usage patterns.

Expected Value

Provide data support for students to optimize booking choices and for venue managers to improve resource allocation.

Macau University Sports Facilities Analysis Workflow



Data Collection

Per analysis needs, collect booking data, venue API info and academic calendar data



Data Preprocessing and Analysis

Convert data format, match code meanings, clarify status labels, complete cleaning and initial screening



Total court analysis

Analyze venue usage volume, time/season/term differences, focus on three key venues

This three-step workflow forms the foundation against which the University of Macau's sports facilities usage optimization recommendations will be developed.

Data Collection consists of the following steps

- 👉 We downloaded sports facility usage data stored in JSON file format via the University of Macau's official data platform (URL: <https://data.um.edu.mo/>).
- 👤 The data encompasses the following fields: bookingID, bookingDate, placeCode, timeSlot, bookingStatus, and lastModifiedDate. Among these, the codes in the placeCode field are encoded to denote venue types; for instance, 'HBC' corresponds to a badminton court.
- 🌐 To clarify the specific names of the venues, we scraped the API information of Sports Facilities from the University of Macau's data platform. As illustrated in the code "sports venue scraper", we obtained the venue information table titled sports_Facilities.
- 🕒 To analyze the correlation between sports venue utilization and university-specific time periods (e.g., final exam weeks, winter and summer vacations), we organized the University of Macau's academic calendar data for 2022-2024 into a table named um academic calendar.

```
DataFrame shape: (611253, 7)

First 5 rows of data:
   _id bookingID          bookingDate placeCode \
0  63b1380f87456b94c60a393d  83943689  2022-12-31T16:00:00.000Z    HSQ1
1  63b1380f87456b94c60a3949  83955369  2022-12-31T16:00:00.000Z  HSQ1~3
2  62d692bd87456b94c67b7075  84083857  2022-12-31T16:00:00.000Z  HTEN5
3  62d6928787456b94c677000f  83943701  2022-12-31T16:00:00.000Z    HSQ1
4  62d6928887456b94c677237c  83955381  2022-12-31T16:00:00.000Z  HSQ1~3

      timeSlot bookingStatus       lastModifiedDate
0            11             0  2023-01-01T02:53:48.000Z
1            11             P  2023-01-01T02:53:48.000Z
2            15             0  2023-01-01T06:15:27.000Z
3            17             0  2023-01-01T06:32:35.000Z
4            17             P  2023-01-01T06:32:35.000Z
```

Figure 1: Content after converting the JSON file

- + Our data was downloaded from the school's website. The original files were in JSON format, so we first converted them into CSV files and saved them for easier access and management in subsequent steps, as shown in Figure 1.

	A	B	C	D	
1	place_code	description	description_chin		
2	HTT9	Table Tennis Court 9			
3	HTEN4_6	Tennis Courts 4_6			
4	HTEN1_3	Tennis Courts 1_3			
5	HGYM	Fitness Room			
6	HTTR	Table Tennis Room			
7	HPAV	Sports Pavilion			
8	HTEN7	Tennis Court 7			
9	HTEN6	Tennis Court 6			
10	HDR	Dance Room			
11	HMPR	Multi-Purpose Room			
12	HTEN5	Tennis Court 5			
13	HTEN9	Tennis Court 9			
14	HTEN8	Tennis Court 8			

Figure 2: Shows the specific meanings represented by partial codes in "sport_facilities.csv".

- ✖ We noticed that in the "placeCode" column, there are many codes with unclear meanings, such as 'BA10', 'BA11', and 'BA12' (not shown in the figure), all of which represent badminton courts. By referencing another file from the school, "sport_facilities.csv", as shown in Figure 2, we identified the specific meanings of these codes and consolidated all venues of the same type but with different codes into a single category.

Sports Facilities Booking Data Dictionary			
Field	Type	Indexed	Description
bookingID	date	Yes	Unique ID of sports facilities booking records
bookingDate	date	Yes	Date of booking, in UTC+08:00 (inclusive)
placeCode	string	Yes	Identity code of sports facilities
timeSlot	integer	Yes	1 – 24 represents 24 hours in a day
bookingStatus	string	Yes	Status of booking: A, U, R, H, O, X, F, P, or N
			Available (A) The time slot is available for booking.
			Free Use (U) If the time slot is not booked, eligible use time slot has been marked for usages, th use.
			Reserved (R) The time slot has been reserved for some
			Hold (H) The time slot has been held.
			Occupied (O) The time slot has been booked.
			Closed (X) The time slot is not open for booking.
			Not Available Due to Dependency (F or P) The time slot is not available for booking

Figure 3: Specific meanings of bookingStatus codes

- + In Figure 1, there is also a column named "bookingStatus" containing single English letters whose meanings were initially unclear to us. According to the instructions on the school's official website, as shown in Figure 3, we identified their specific meanings, thereby preparing adequately for subsequent analysis and visualization.

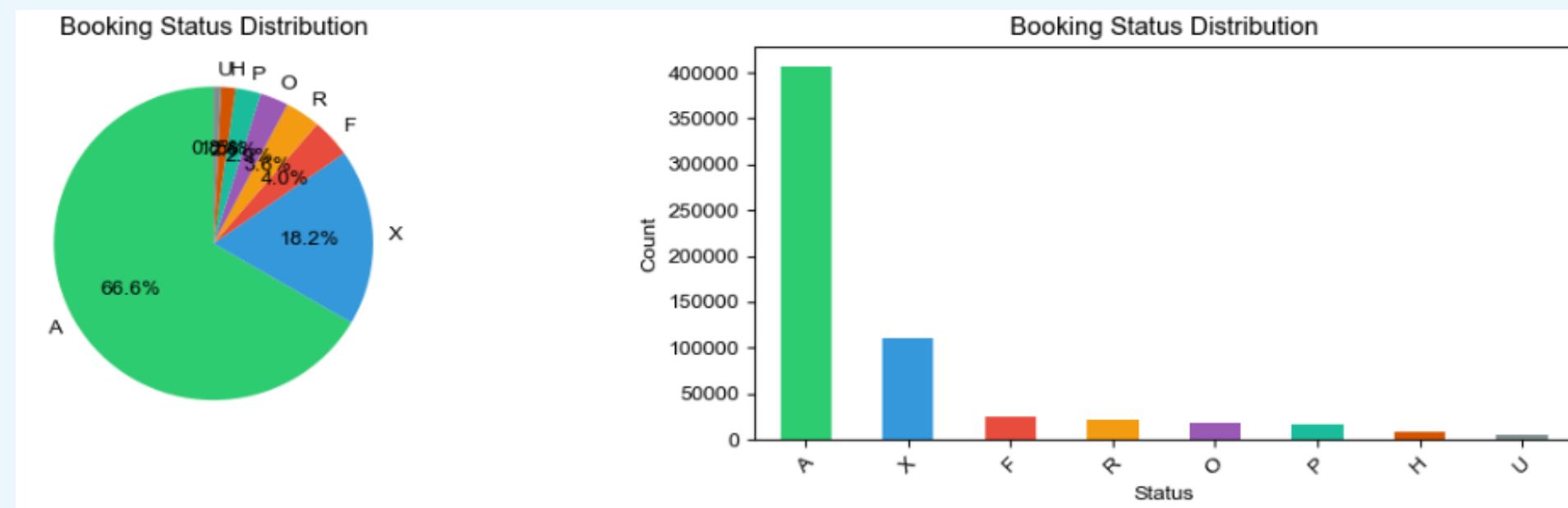


Figure 4: Distribution of various codes in bookingStatus

- ✖ In Figure 4, we present the distribution of these codes within our dataset. It can be observed that "**available**" accounts for the vast majority of the data, indicating that a large number of venues were not booked.
- ✚ Finally, after confirming there were no missing values, we organized our initial dataset and created an initial visualization. Figure 5 shows the total number of bookings for all venues in our records, aggregated from December 2022 to January 2024.

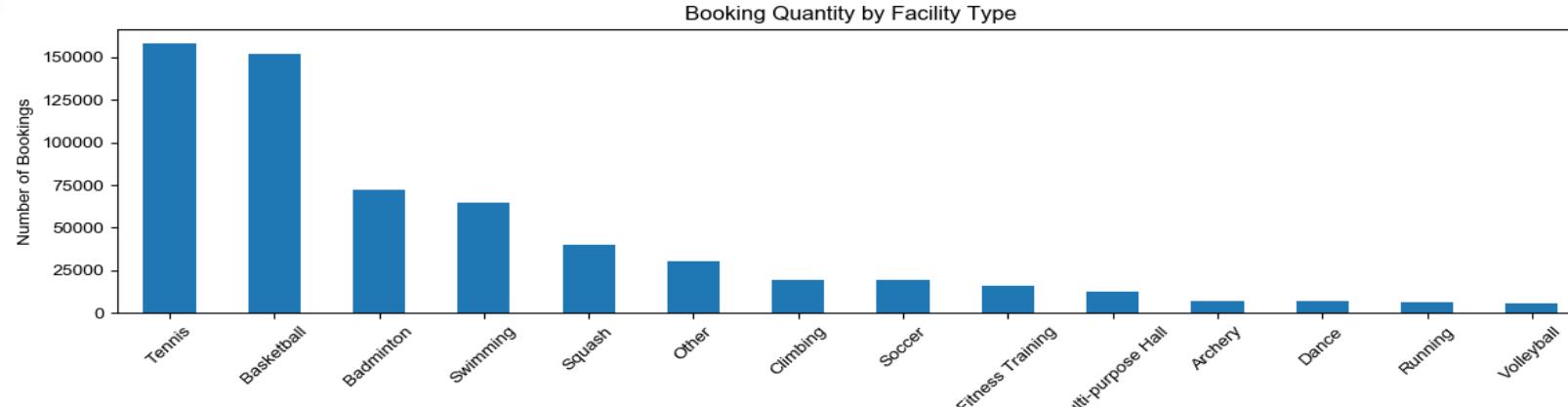


Figure 5: Total number of bookings per venue

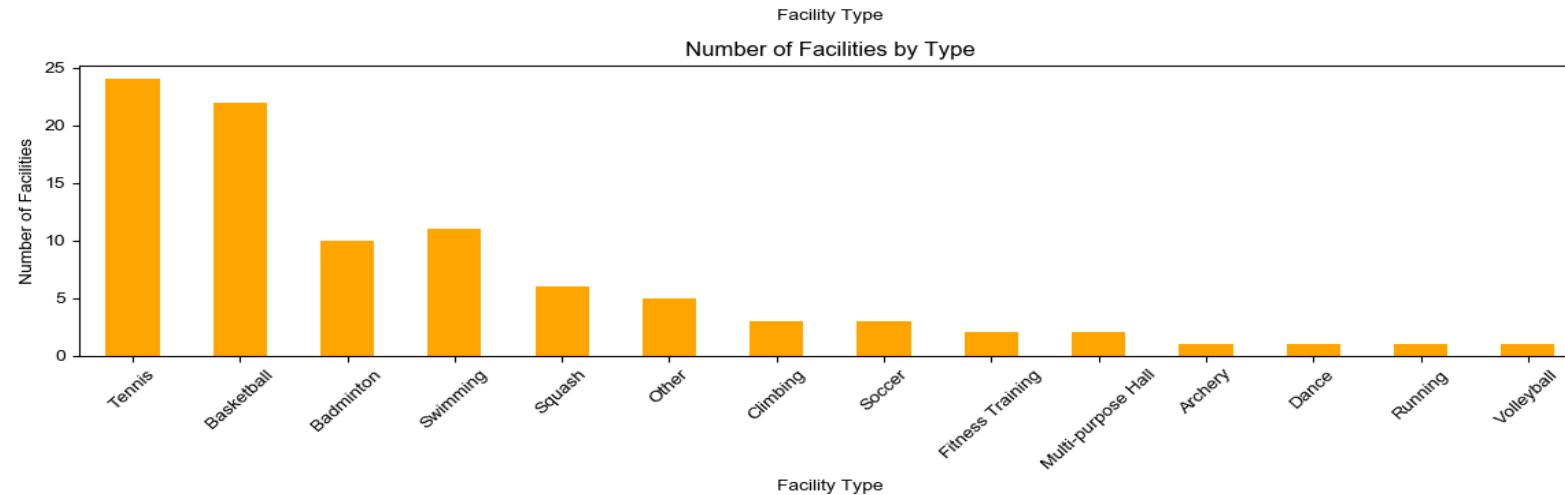


Figure 6: Quantity of each venue at the school

- + Figure 6 show the number of each venue at our school.
- ✗ To clearly illustrate the relationship between Figure 5 and Figure 6, I used a treemap to visualize their connection.

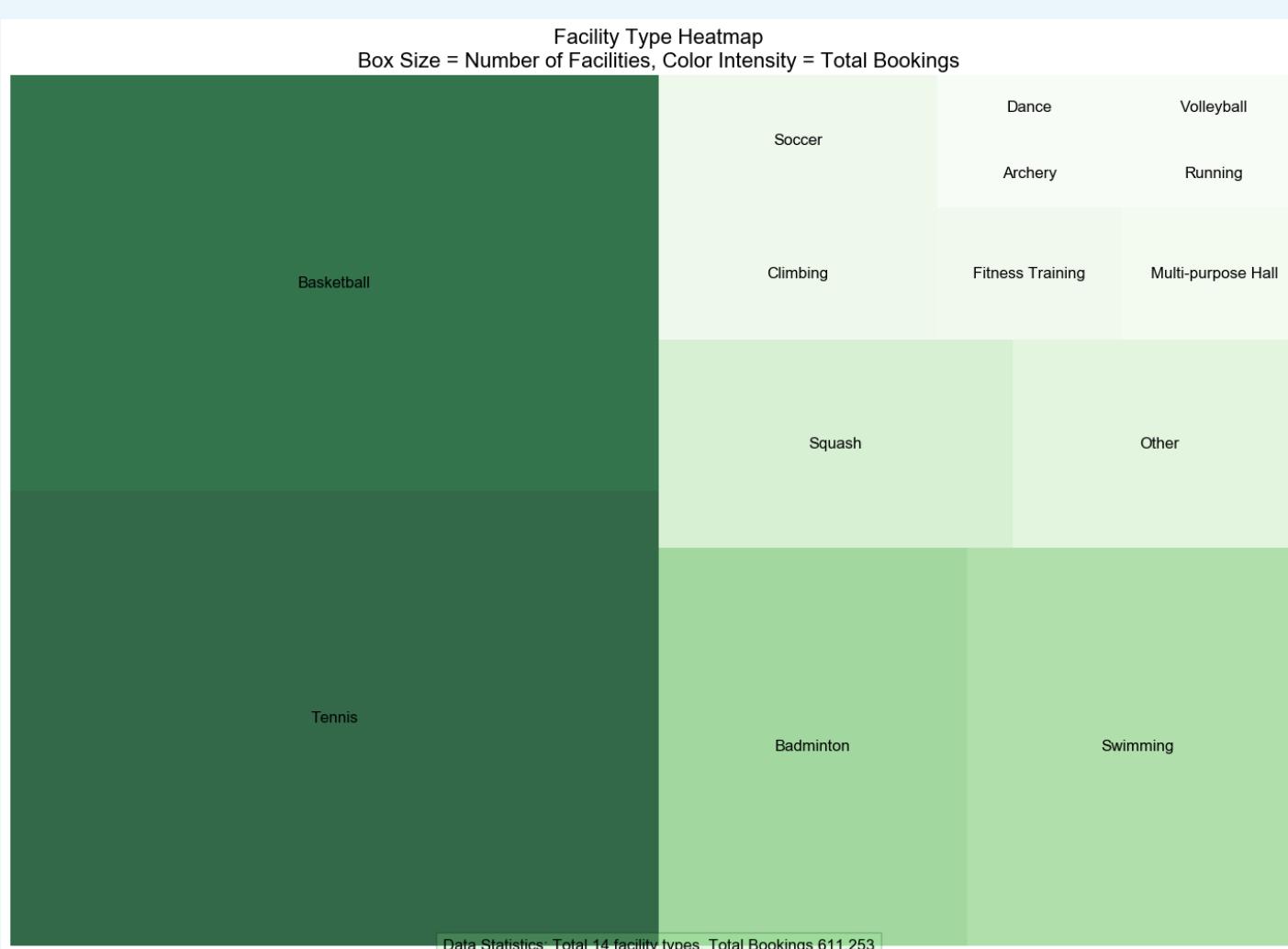


Figure 7: Relationship between the number of bookings and the quantity of venues

- + In Figure 7, the size of each box represents the quantity of venues, while the color intensity indicates the level of usage (number of users).
- ✗ From this chart, we can intuitively see that **tennis and basketball are the most frequently used sports activities at our school**, and they also have the largest number of dedicated venues.

Total court analysis

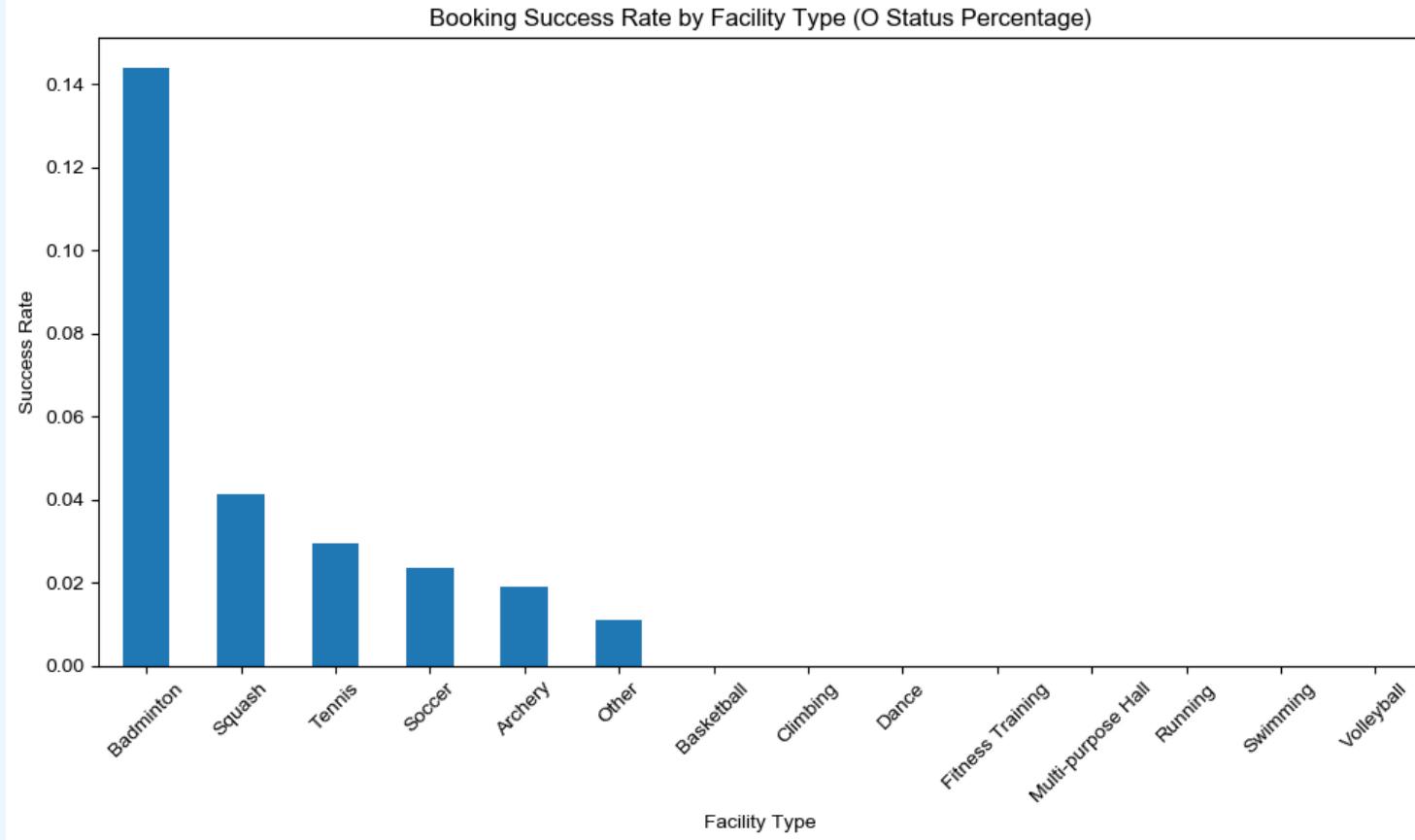


Figure 8: Quantity of venues marked with status "o"

- + Our goal is to identify the optimal times for using sports facilities, so we analyzed data based on the "o" status (representing successful bookings). As shown in Figure 8, badminton has the highest usage, but popular venues like basketball courts lack "o" status records due to unsynced offline registration data. Subsequent analysis will focus on three representative venues:
Badminton, Tennis, and Squash.

- + Now that we have identified which venues to analyze, examining only these three is insufficient for providing constructive recommendations.
- Therefore, our second focus is **time**. The dataset includes usage data for each day from 7:00 to 22:00 over nearly two years.

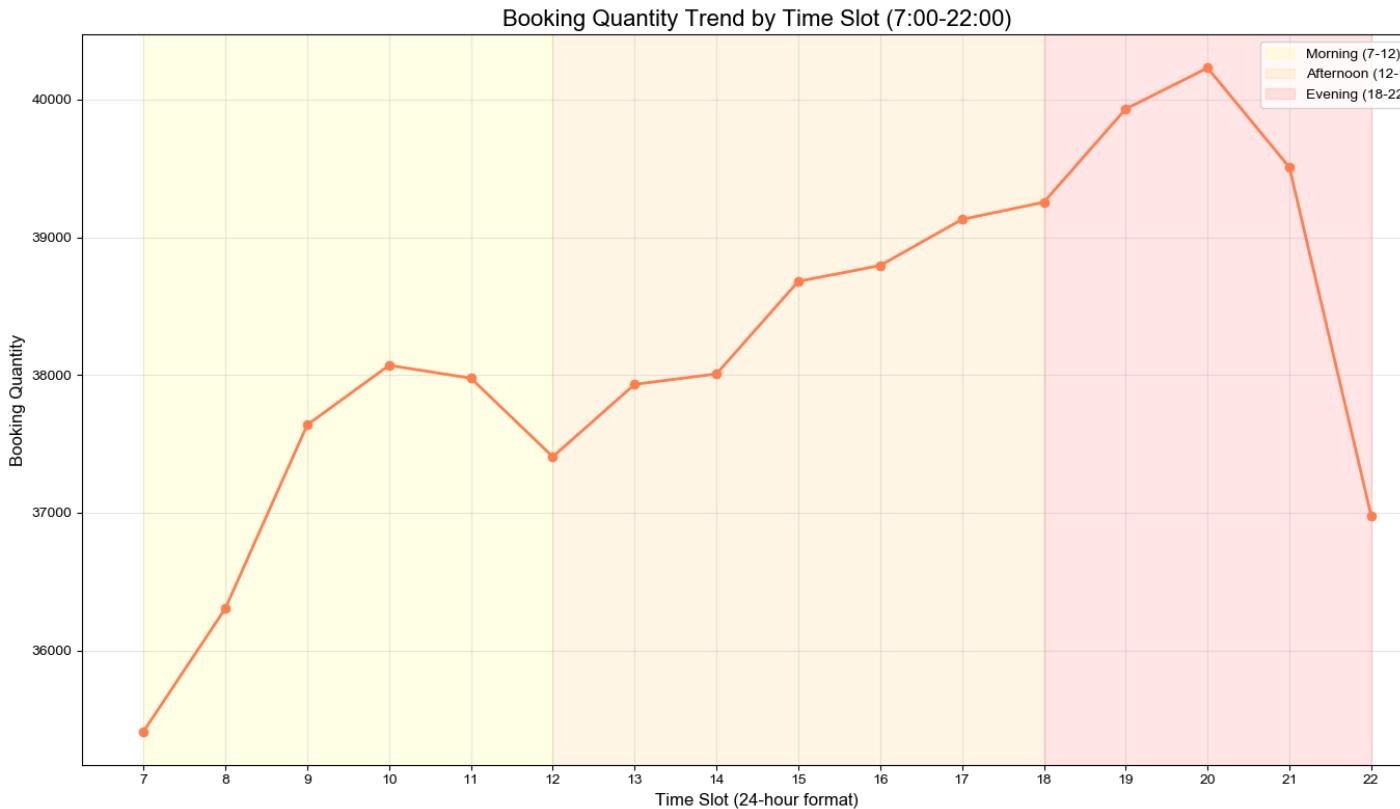


Figure 9: Usage of all venues across different time slots

- ✖ First, we observed the daily usage patterns. In Figure 9, we can see that usage steadily increases, peaking at 20:00 (8 PM). This indicates that **most people prefer exercising in the evening**.

Total court analysis

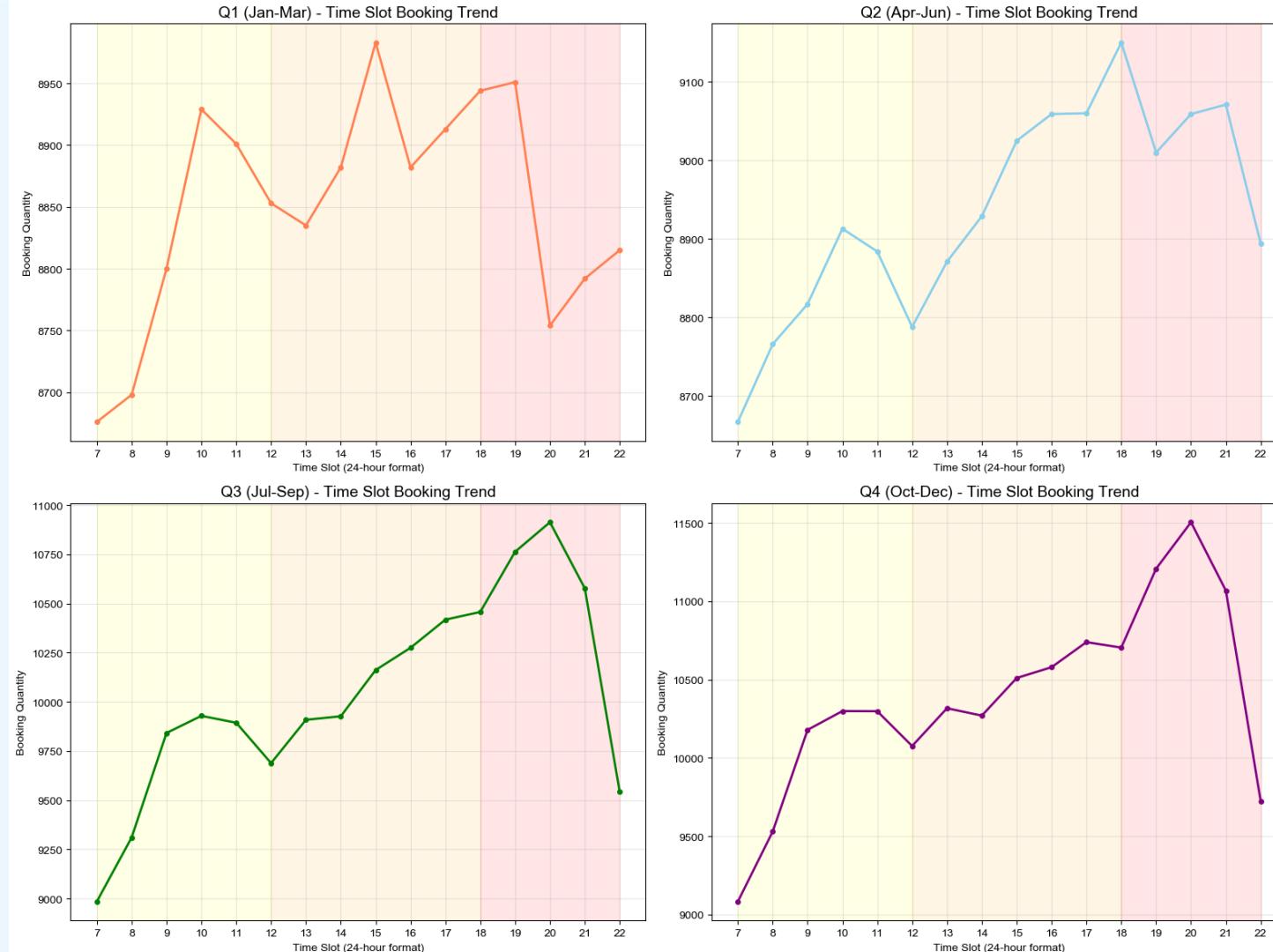


Figure 10: Line chart of user count for each season

- + Next, we analyzed usage patterns based on different seasons. In Figure 10, most trends are similar to those in Figure 9. We noticed significant fluctuations during the spring season. Based on our analysis, we believe **holidays are the primary reason for these substantial fluctuations**, as this period includes the Christmas holiday and Chinese New Year.
- ✗ Therefore, we divided the time periods into **academic terms and holiday** periods based on the school calendar to observe whether the expected significant fluctuations appear in our data.

- As shown in Figure 11, our hypothesis is confirmed: the periods during **school terms and holidays** indeed affect the usage of our school's sports facilities.

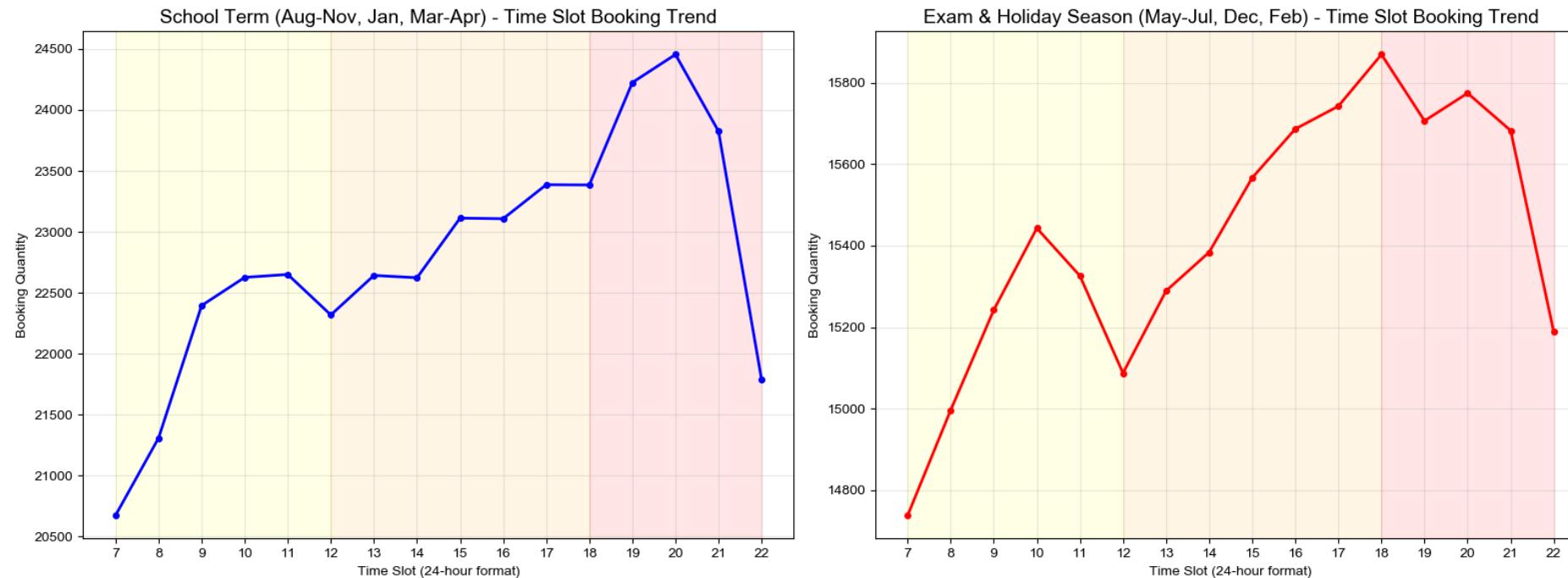
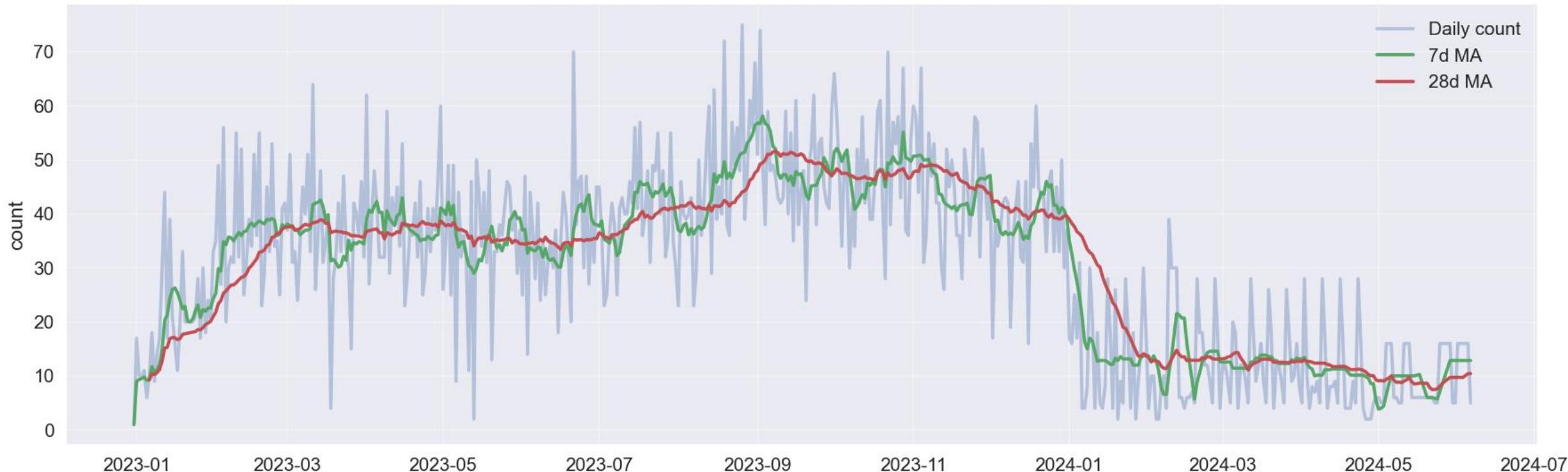


Figure 11: Line chart of usage after differentiating between academic terms and holiday periods based on the school calendar

- Therefore, based on all the factors mentioned above, we will focus on analyzing the **top three most frequently** used venues and provide recommendations accordingly.



Badminton court daily occupied count with 7d/28d MAs



Data Extraction

Established a dedicated badminton dataset by extracting records associated with "HBC" venue codes

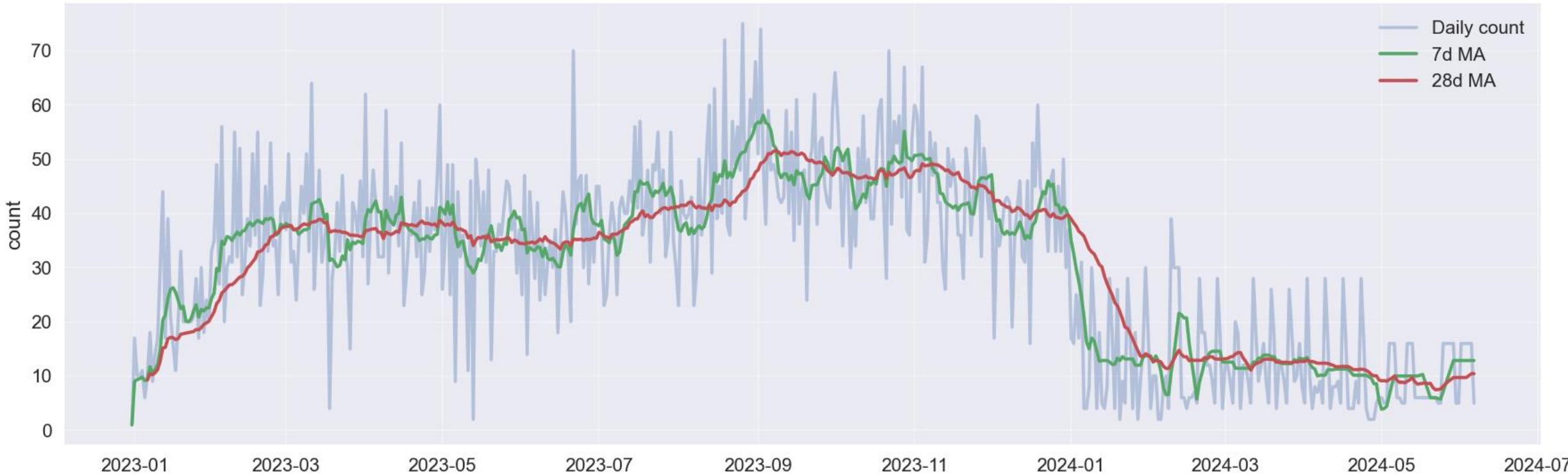
Booking Status Distribution

- Dominant Statuses: Identified X and A as the most frequent categories.
- Data Grouping: Merged O, H, and R into "Occupied Status Z" for clearer unavailability analysis.

Data Processing and Time Series Construction

Standardized the dataset into a daily time series and addressed data sparsity by generating a continuous date range, utilizing forward fill imputation for missing

Badminton court daily occupied count with 7d/28d MAs



⌚ Time Series Analysis

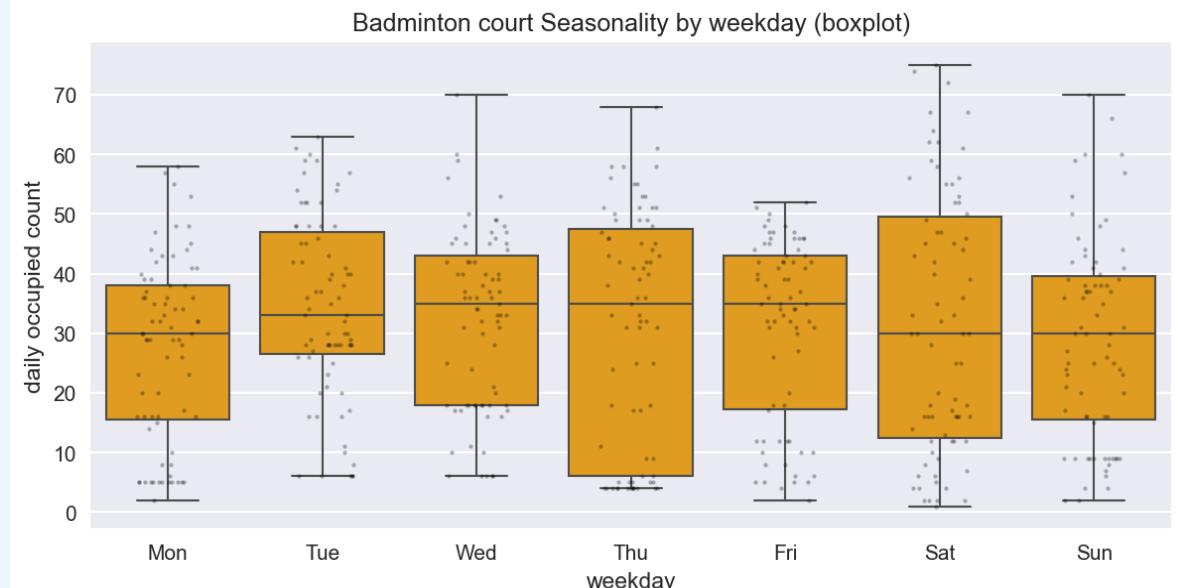
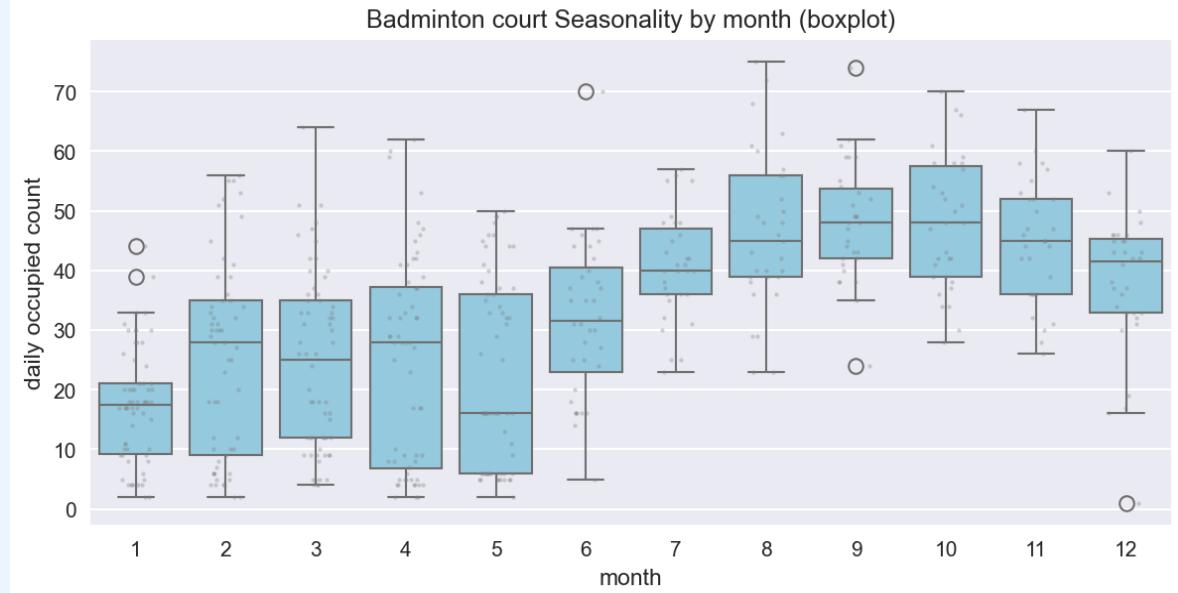
Visualized temporal trends by superimposing 7-day and 28-day moving averages on the daily occupancy curve to distinguish short-term fluctuations from long-term patterns.



Seasonal Analysis

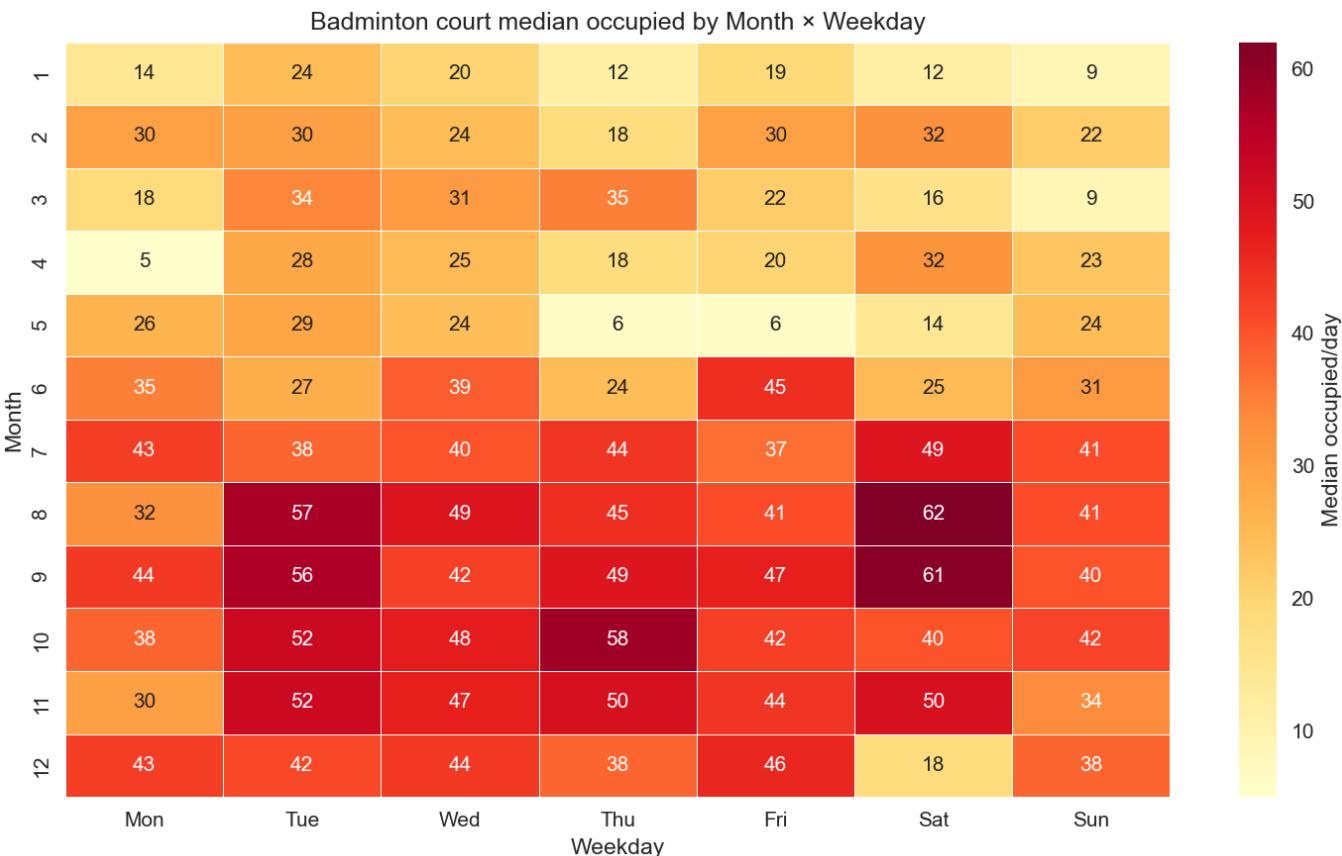
Why Mangkut (2018) Validates Our Findings

- Monthly Trend: Box plot analysis revealed a significant increase in average daily usage from August to October.
- Weekly Pattern: No major difference was observed in median usage between weekdays and weekends.





Seasonal Analysis

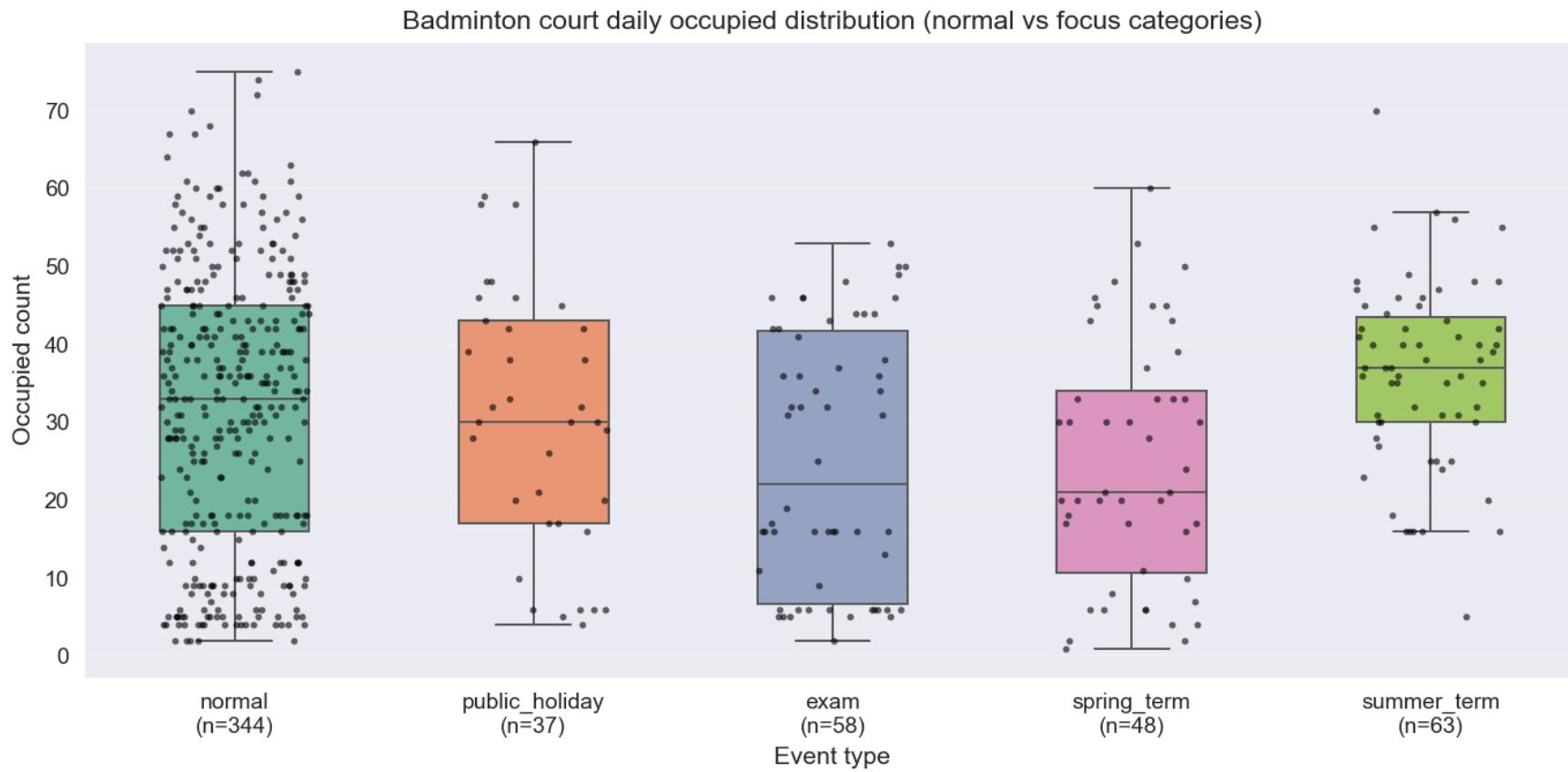


Heatmap analysis revealed usage patterns driven by three key factors:

- Academic Calendar: Surge in August due to the start of the new semester.
- Weather: Preference for indoor badminton over outdoor sports (e.g., tennis) during high-heat/UV summer months.
- Holidays: Notable dip in usage during Q1 festive periods (Jan–Mar).

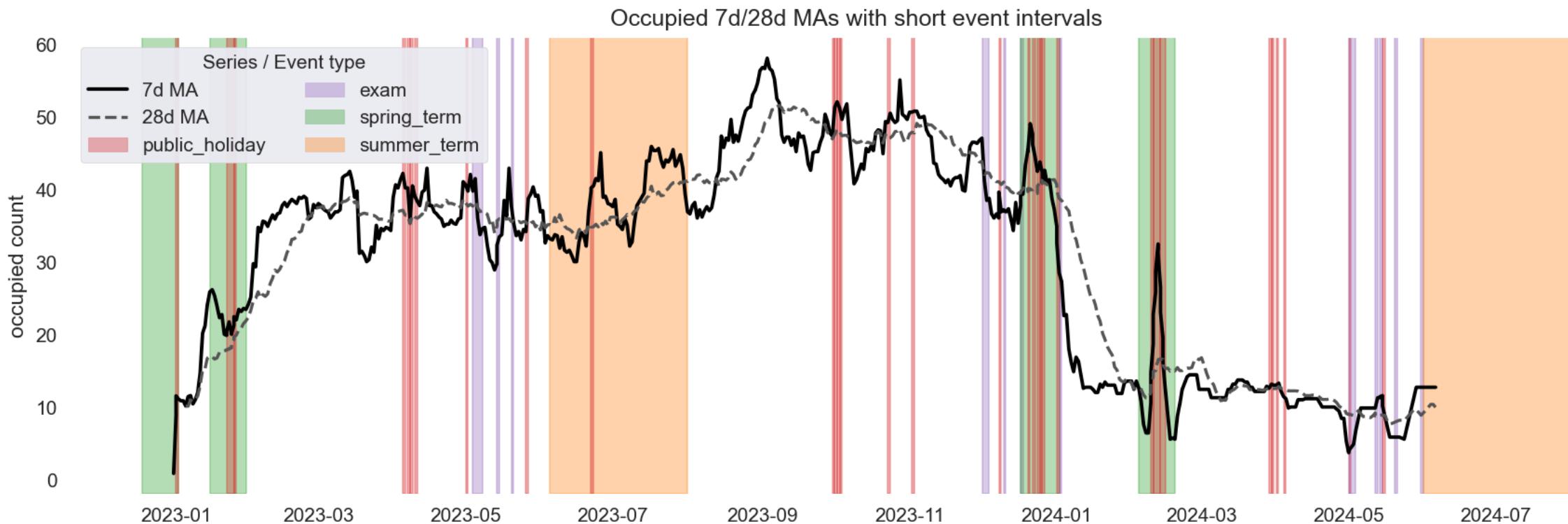


Academic Calendar Correlation



Impact of Academic Schedule on Usage:

- Exams & Winter Break: Identified as the primary drivers for low occupancy (Significant Drop).
- Summer Term: Correlates with a slight increase in facility utilization.
- Public Holidays: Surprisingly, these showed no significant deviation from normal usage patterns



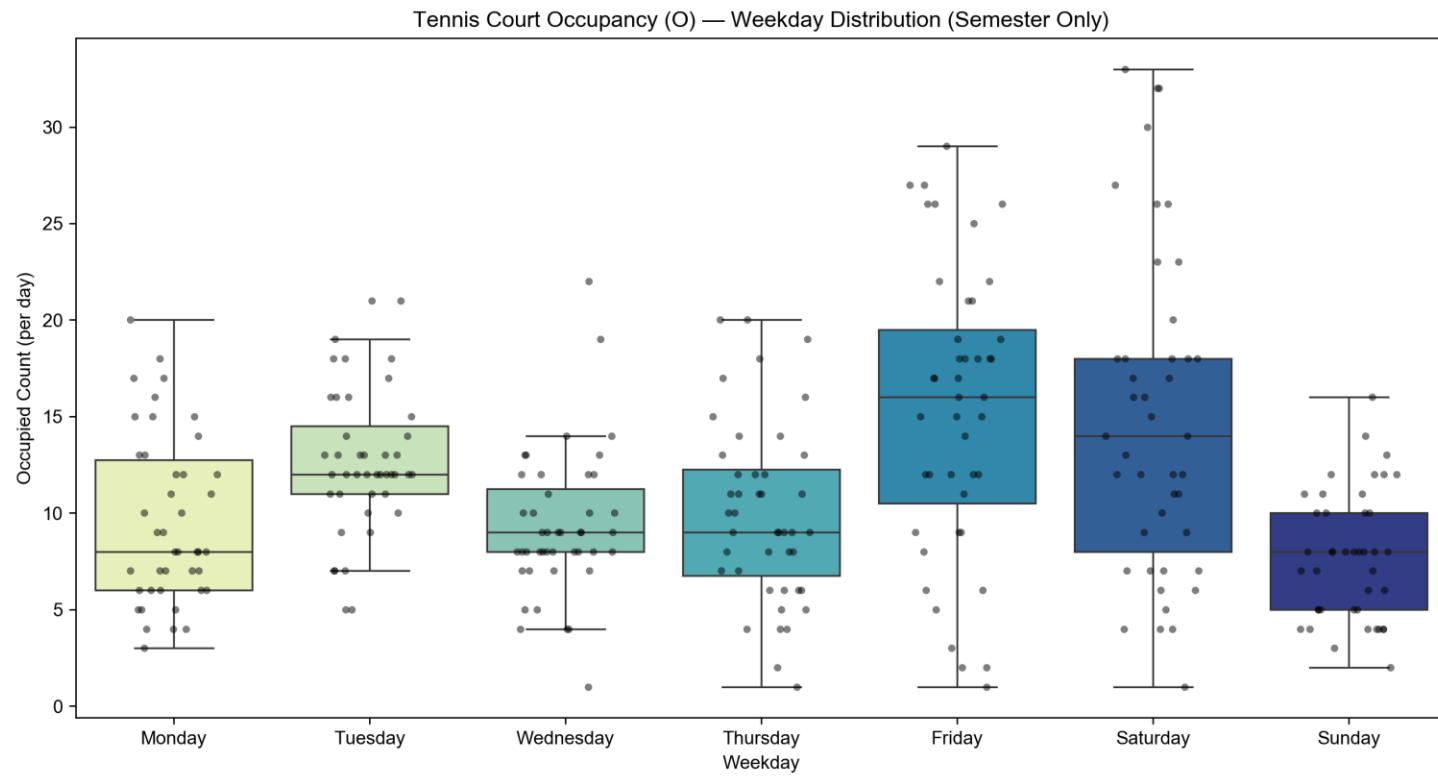
⌚ Time Series Analysis

Visualized temporal trends by superimposing 7-day and 28-day moving averages on the daily occupancy curve to distinguish short-term fluctuations from long-term patterns.



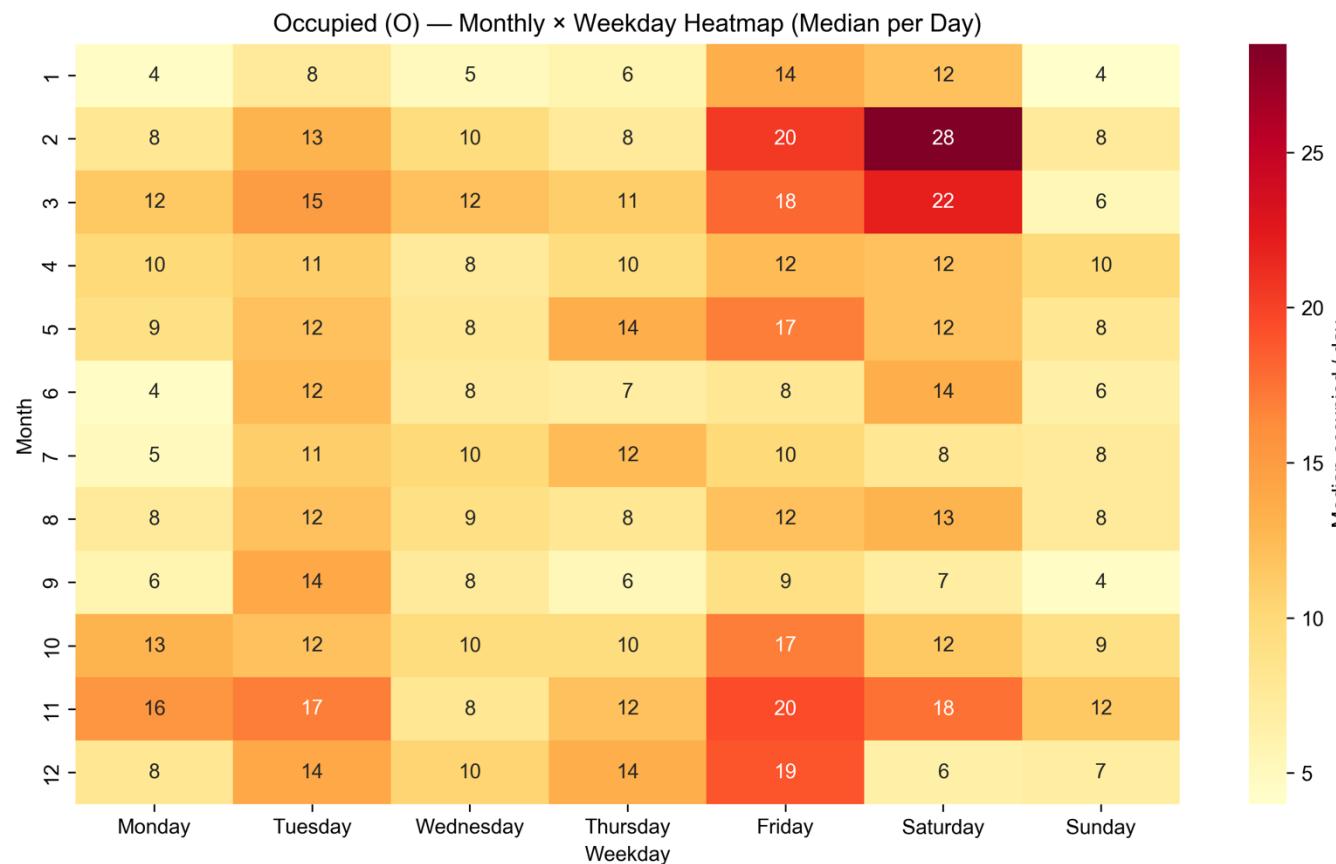
Weekly Distribution Analysis

- Peak Activity: Fridays and Saturdays show the highest and most variable usage.
- Weekly Trend: Usage is stable from Monday to Thursday but drops relatively low on Sundays.



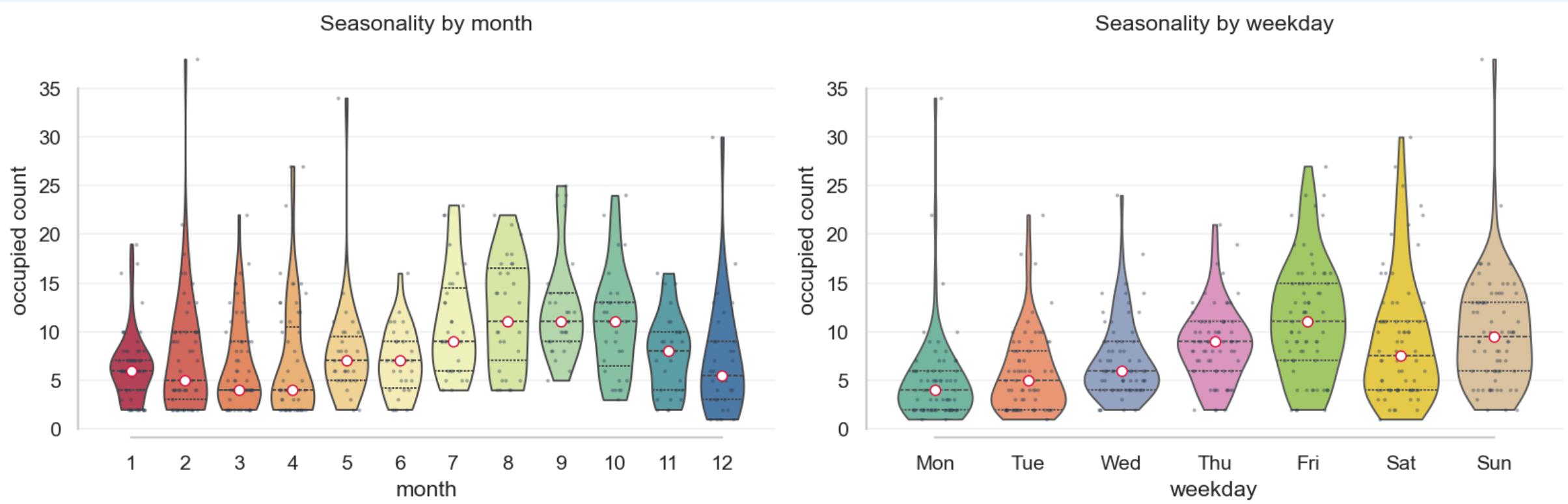


Seasonal Analysis



Tennis usage is heavily influenced by seasonal climate conditions:

- Spring Peak (Feb–Apr): Highest usage observed, particularly on Fridays and Saturdays, driven by favorable outdoor weather.
- Summer Dip (Jun–Sep): Activity drops significantly due to high temperatures.
- Weekly Pattern: Sundays consistently show lower engagement regardless of the season

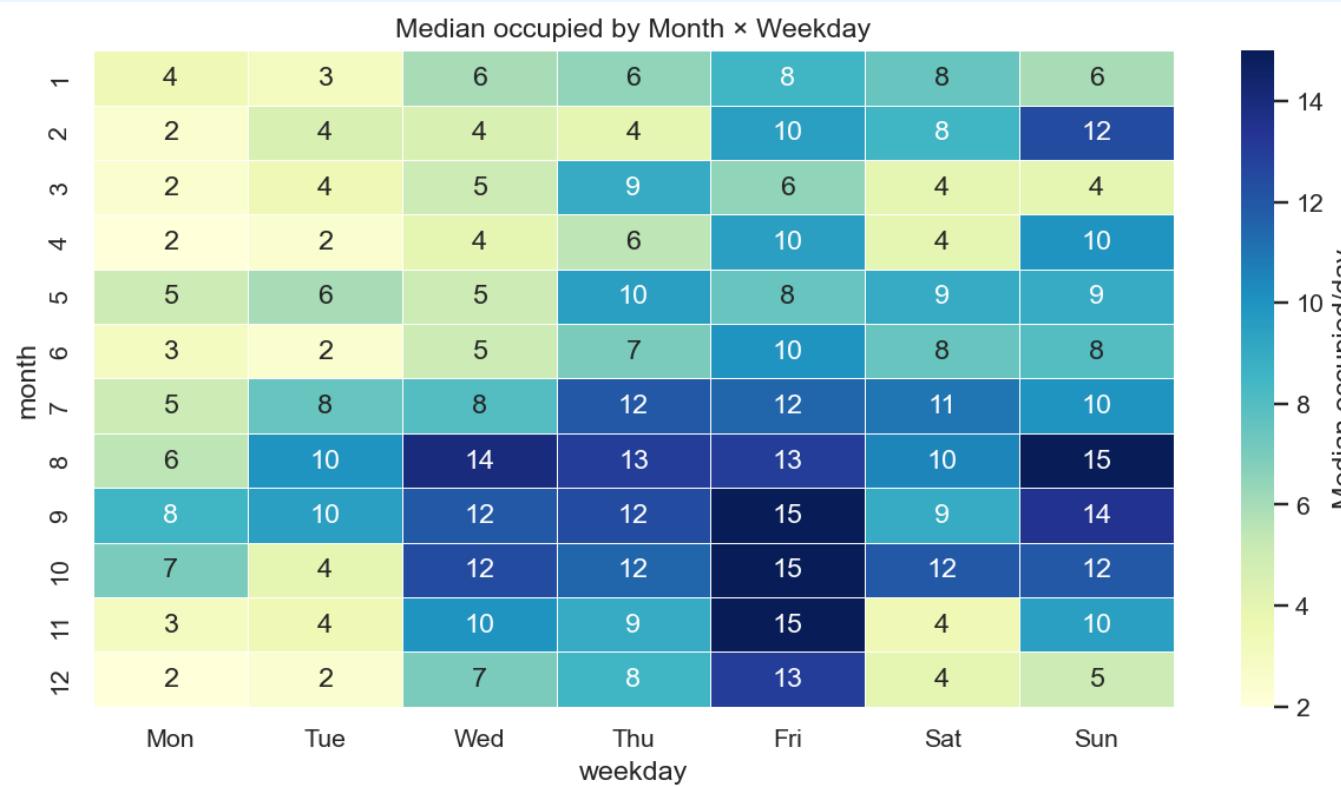


Overview

- Seasonal Trend: Usage peaks slightly from August to October, with dips observed early in the year and during holidays.
- Weekly Pattern: Unlike the consistent daily demand for badminton, squash shows a clear preference for Fridays and weekends, indicated by wider usage ranges.



Seasonal Analysis

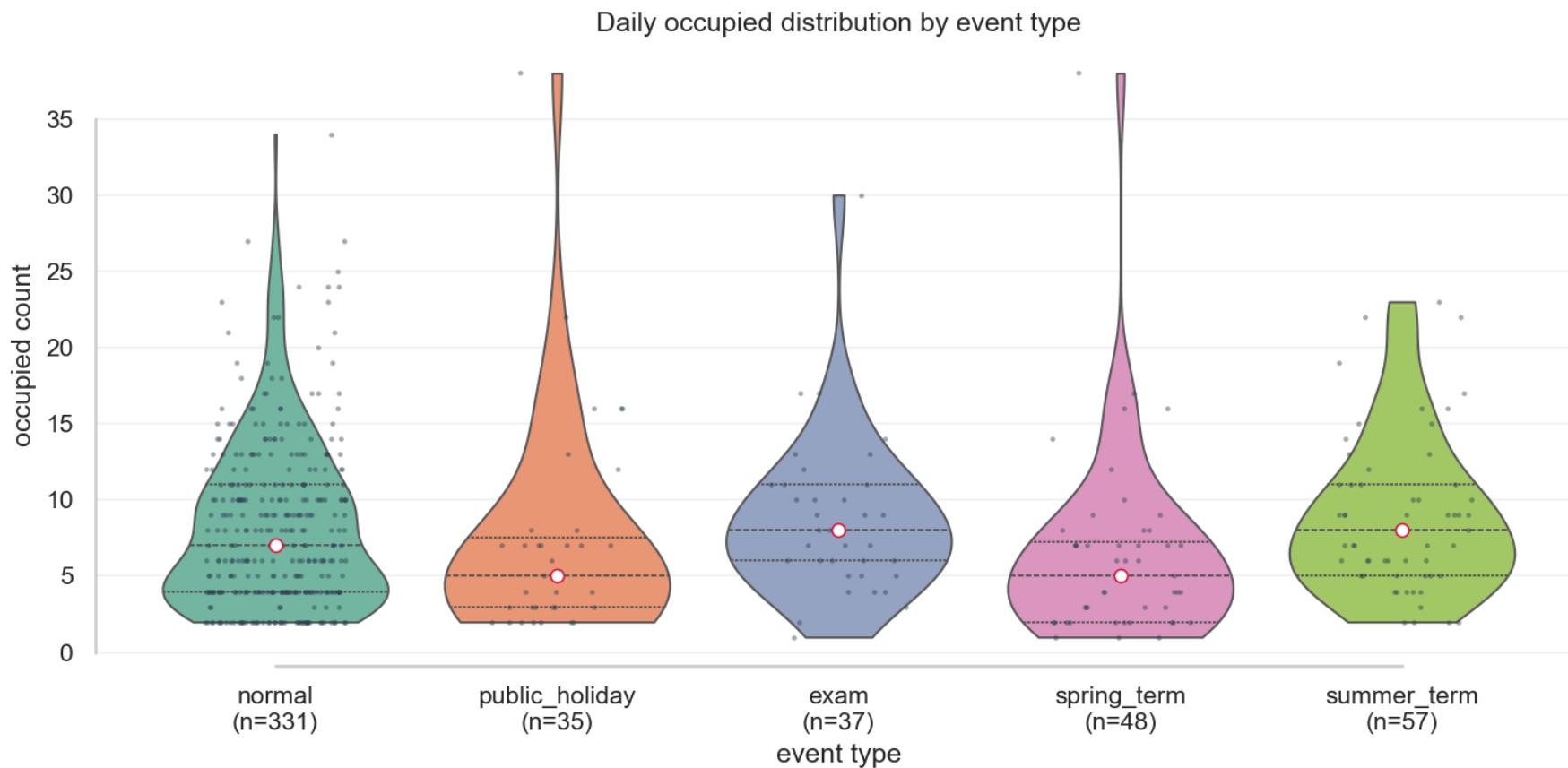


Similar trends observed for both indoor facilities:

- Summer Term: Highest median usage (Peak Demand).
- Normal School Days: Widest variability in usage.
- Exam Periods: Lowest and most stable activity, reflecting high academic pressure.
- Public Holidays: Lowest median levels due to reduced campus presence.



Comparison Across Holiday, Exam, and Term Periods



Breakdown of Daily Occupancy by Event Type:

- Summer Term: Demonstrates the highest median usage (≈ 8) and consistent engagement.
- Normal Days: Exhibits the widest variability, ranging from quiet days to peak occupancy exceeding 30.
- Low Activity: Public Holidays and Spring Term show the lowest median usage (≈ 5), indicating reduced interest during these periods.
- Exam Period: Surprisingly stable usage with a relatively high median, suggesting sustained participation despite academic pressure.

Machine Learning Model Construction & Performance Evaluation

- Label Redefinition: To address data sparsity in the "O" label, we redefined "A" and "U" records as positive samples to model "Bookability" effectively.
- Feature Engineering: Focused on 4 facility types (Tennis, Basketball, Badminton, Squash) and integrated multidimensional features including Time Slots, Weekday/Weekend, and Seasonality.
- Model Selection: The Random Forest model achieved 79.42% accuracy, significantly outperforming Logistic Regression (74.48%) and demonstrating superior precision in identifying successful bookings.

```
Loading dataset...
Dataset basic information:
Data shape: (422448, 7)
Facility type distribution:
facility_type
Tennis      158254
Basketball   152164
Badminton    72108
Squash       39922
Name: count, dtype: int64
Booking success rate: 66.54%

Performing feature engineering...
Feature engineering completed!
Number of features after processing: 18
Training set size: (337958, 15)
Test set size: (84490, 15)

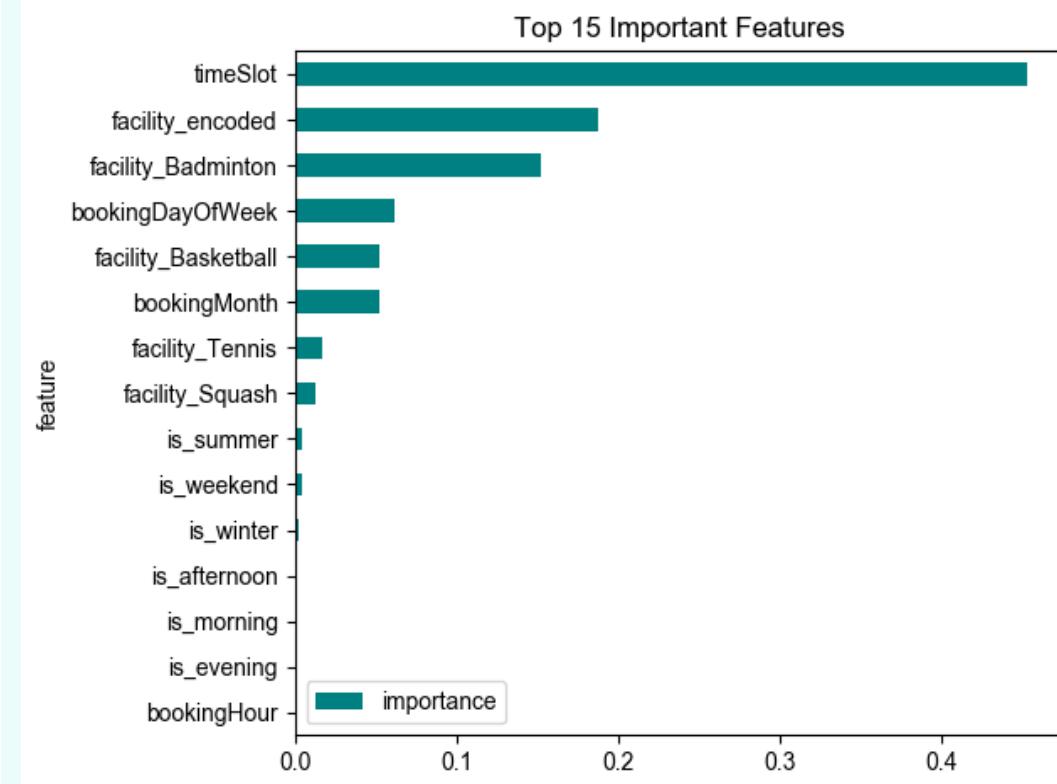
Training Logistic Regression...
Logistic Regression accuracy: 0.7448

Training Random Forest...
Random Forest accuracy: 0.7942

Best model: Random Forest
...
accuracy           0.79          0.79        84490
macro avg         0.77          0.75          0.76        84490
weighted avg      0.79          0.79          0.79        84490
```

Analysis of Key Drivers for Booking Success

- Time is Critical: Feature importance analysis reveals timeSlot as the most dominant factor. When you book matters more than what you book.
- Facility Variance: facility_encoded and specific sports (e.g., Badminton, Basketball) are top-tier features, indicating significant variance in competition for different venues.
- Cyclical Patterns: DayOfWeek and Month contribute meaningfully to the prediction, whereas seasonal factors (Summer/Winter) show lower impact.





==== Business Insights and Recommendations ====

Easiest facility to book: Basketball (success rate: 79.69%)

Best booking time: 16:00 (success rate: 66.54%)

Weekday success rate: 67.21%

Weekend success rate: 64.86%

==== Prediction Examples ====

Tennis - Sat 18:00 - Booking success probability: 78.27%

Badminton - Wed 10:00 - Booking success probability: 38.75%

Basketball - Sun 20:00 - Booking success probability: 95.48%

Project completed! You can provide data-driven suggestions for school facility management based on this analysis.

🔗 Business Insights & Intelligent Booking Recommendations

- Macro Insights: Easiest facility to book: Basketball (~79.7%); Best booking time: 16:00 (~66.5%).
- Real-time Prediction:
 - High Success: Sun 20:00, Basketball -> 95.48%
 - Low Success: Wed 10:00, Badminton -> 38.75%
- Practical Value: Provides data-driven guides for students to choose off-peak times and supports management in optimizing resource allocation.



Key Findings

- Key Drivers: Sports facility usage at the University of Macau is strongly influenced by time of day, season, and the academic calendar.
- Usage Patterns: Evenings are consistently the busiest, while activity significantly drops during exams and holidays.
- Environmental Factors: Indoor sports show stability across seasons, whereas outdoor sports depend heavily on weather conditions.



Strategic Recommendations

- For Students: Try to avoid peak hours where possible based on data trends.
- For Managers: Leverage data and prediction models to improve scheduling and resource allocation.



Project Value

- Overall, this project demonstrates how data analysis can enhance campus life, making sports facilities more efficient and user-friendly.

Reference:

1. Ortigossa, E. S., Dias, F. F., Nascimento, D. C., & Nonato, L. G. “*Time Series Information Visualization –A Review of Approaches and Tools,*”arXiv (2025). . <https://arxiv.org/pdf/2507.14920v1>
2. Endo, R., & Hosobe, H. “*Visualization of Time Series Data Using Clustered Heatmaps and Line Graphs,*”VINCI 2024, ACM (2024).. <https://dl.acm.org/doi/fullHtml/10.1145/3678698.3678705>
3. Luohen de Hanjia, Data Analysis & Visualization Python Map Plotting 3 —GeoPandas Usage Essentials,” CNBlogs (n.d.).. <https://www.cnblogs.com/luohenyueji/p/17485432.html>
4. Ali Yiyang. (n.d.). Python: *Using Pyecharts to Draw China's Population Distribution and Migration Maps..* <https://download.csdn.net/blog/column/10599192/113864429>
5. Kairan Yang ,Yujun Xie ,Hengtao Guo. *Optimization of spatial distribution of sports parks based on accessibility analysis,*”PLOS ONE (2023).. <https://doi.org/10.1371/journal.pone.0291235>



澳門大學
UNIVERSIDADE DE MACAU
UNIVERSITY OF MACAU

Thank you!

Group Members: Liang Zhongkai He Jingping Zhang Yi Liang Zhengping

2025.10.14