The UConn Recreation Center Usage Trends:

A Student Led Statistical Anlysis

Emmet Spaeth
University of Connecticut
emmet.spaeth@uconn.edu

David Matos
University of Connecticut
david.matos@uconn.edu

Abstract

The aim of this paper is to determine trends in the UConn rec center usage. This paper will analyze the occupancy of the rec center both over the course of a day, broken into fifteen minute increments, and over a week, comparing each day. Then, we will to explain the trends in the data. Notably, the difference in usage between the days of the week and the periods of growth throughout a day. Finally, we will provide our explanations of these trends, advice to improve the quality of service and potential cost saving measures.

Contents

1. Introduction

When should I go to the rec? This is the driving question of the paper. Many students have shared the experience of having to change a workout because all the machines they need are full or having to work in with a complete stranger. It can be intimidating and uncomfortable for some people to work out when the rec is at its busiest. Equipped with the knowledge of when the rec is busiest, many students feel comfortable going to work out for the first time.

The current solution provided by UConn is through the rec center app. It's a webpage that tells you the how full the rec is as a percent of its total capacity. The webpage acts like a ping, only telling you the capacity at that exact moment. It can't tell you what the capacity will look like in a couple hours when you actually want to go work out. Unless you spend a ton of time paying attention to occupancy on the app, you have no clue which times or days have the least people. The app also can't tell you if it's busier than usual or just a random fluctuation, you just have to go and find out. We set out to get real data to solve these problems.

2. Methods

2.1. Data collection

The only source of information that is publicly available is the capacity page in the rec app. Many of UConn's apps are made by embedding Chromium and using web technologies (HTML, CSS, JavaScript), most likely through the popular Electron framework. This mean once we found the underlying webpage a simple web scraper could record the occupancy.

After finding the web page(https://app.safespace.io/api/display/live-occupancy/86fb9e11) we wrote a script in python to read the value and record it with a time stamp. The script was set to run every fifteen minutes with a cron job on the server.

The information on the web page was recorded into a CSV file with the following format. If the script threw any kind of error while running, it would simply record "Error,Failed to Pull Data at ***DATETIME***". With a successful execution the script would record, in this order, the month, day of the month, day of the week, time, and the data it read (the occupancy).

For example, here is the first five entries:

```
Month,day,weekday,time,occupancy
Oct,10,Fri,18:30:04,408
Oct,10,Fri,18:45:04,450
Oct,10,Fri,19:00:09,469
Oct,10,Fri,19:15:04,471
Oct,10,Fri,19:30:04,486
```

Data was collected starting October tenth and ending December tenth. There was a brief period without any date due to an unexpected outage. Thanksgiving break was another period without usable data leading to a total of nineteen days down during the data collection period.

2.2. Data filtering and cleaning

Our data cleaning pipeline is as follows:

1. Read CSV data using the pandas library.
2. Convert the time and date related columns into a single datetime object.
3. Remove entries where the data was missing or invalid.
4. Adjust the timezone to Eastern Standard Time (EST).
5. Handle the transition during day light's savings time.
6. Filter data from outside of the operating hours of the rec.
7. Group data by day of the week, weekday/weekend, or by hour of day.
8. Visualize and average the pre mentioned groups.

3. Results

The first piece of analysis we did was a graph of the average occupancy for each day over time, shown in Figure 1. It was immediately obvious where the gaps in our data were. However, on first glance our data seemed to be fairly random. The mean occupancy was always lower than the median, suggesting outliers may be bringing the mean down or central values were significantly greater than their surrounding data.

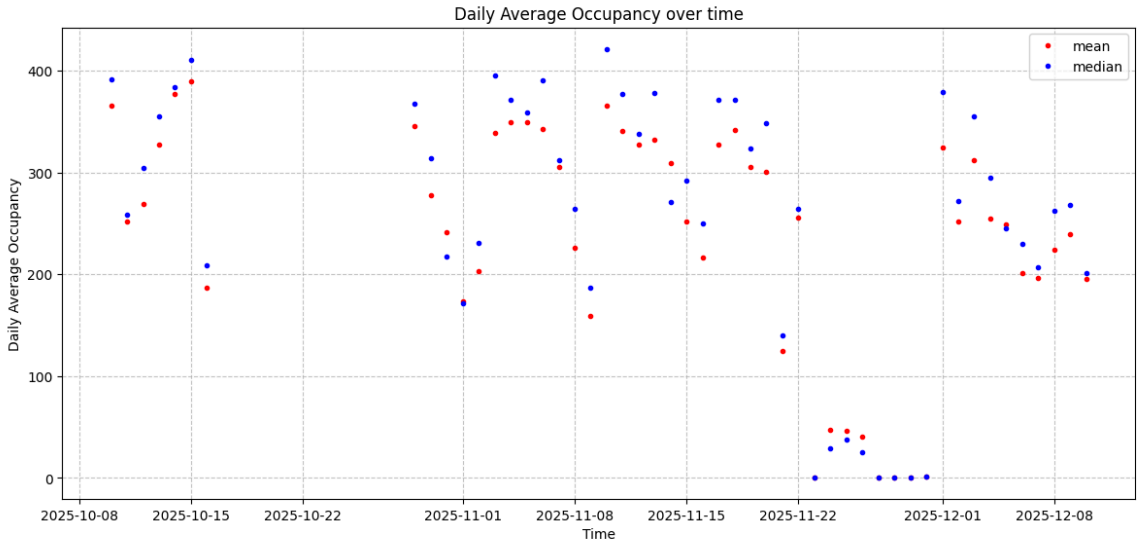

Figure 1: A graph of the daily average occupancy over time

Our next step was a breakdown of the average occupancy by day of the week, shown in Figure 2. This graph gave a lot of insight. First, it told us the most popular day to go to the rec was Wednesdays. Followed closely by Mondays and Tuesdays. After which, there is a steep drop off in usage starting on Thursday with each following day seeing less usage than the last. Sunday is the least popular day to go to the rec.

Looking back at Figure 1 we can actually see this cyclic pattern. It's important to remember the limited hours at the rec center on the weekends are probably contributing to the lower

daily average. However, the reduced hours on the weekend doesn't explain the large difference between Saturday and Sunday.

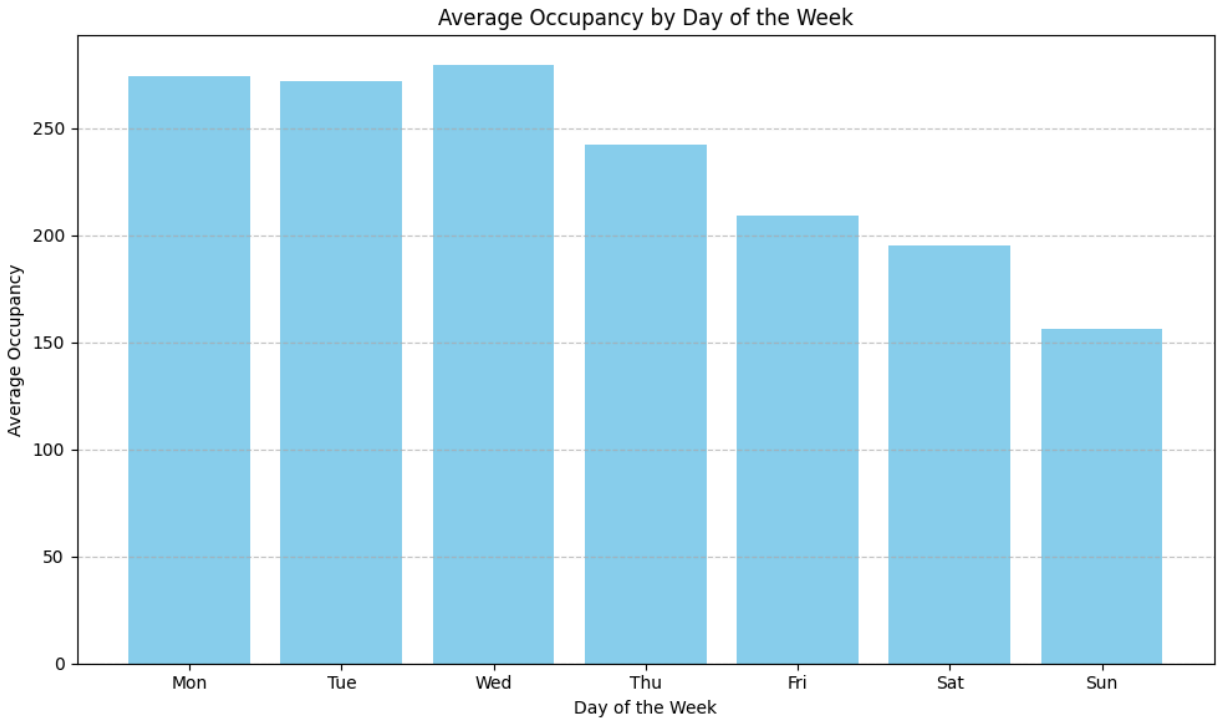

Figure 2: A graph of the daily average occupancy by day of the week

The most interesting of the data is the graph of occupancy over an average day. The peak on both weekdays and weekends is around hour 17 (5pm). This is closing time on the weekend, but on weekdays it means a decrease in total occupancy for a couple hours until closing time.

There are three main periods of growth during a weekday. One from 6-8, the second from 11-13, and the last from 15-17. On the weekend the first two hours of the day 10-12 sees the greatest growth. After that period, growth slows and some what stagnates, but occupancy doesn't decrease until closing.

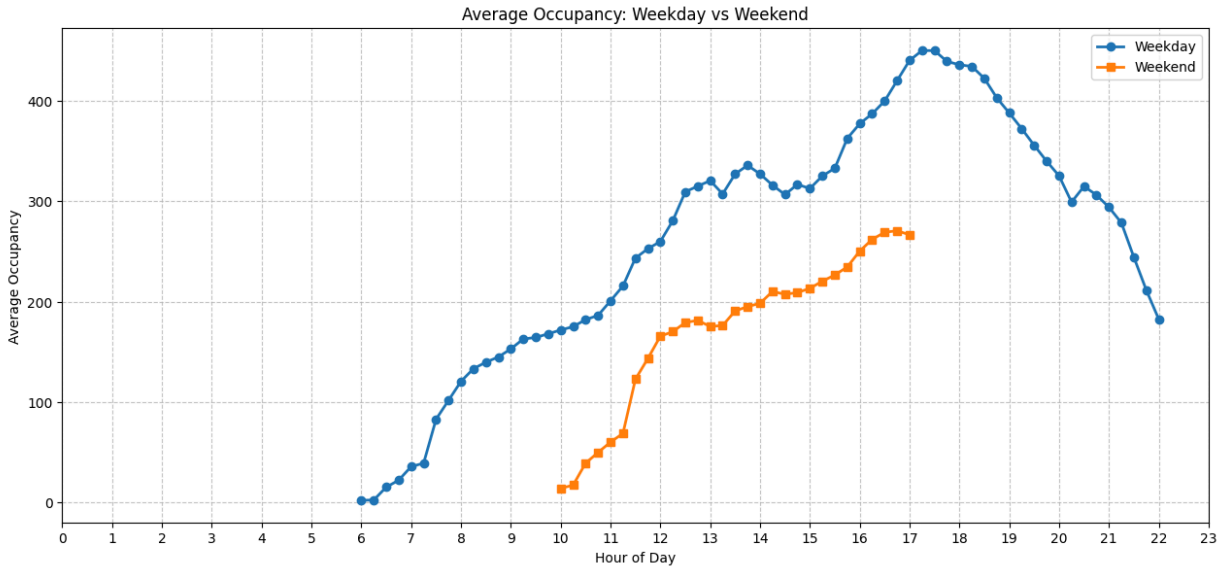

Figure 3: A graph of occupancy over time separated by weekday vs weekend

To further analyze the flow of people in and out of the rec we graphed the first order derivative of the occupancy over time graph in Figure 4. In Figure 4 we see the same periods of growth in the first graph.

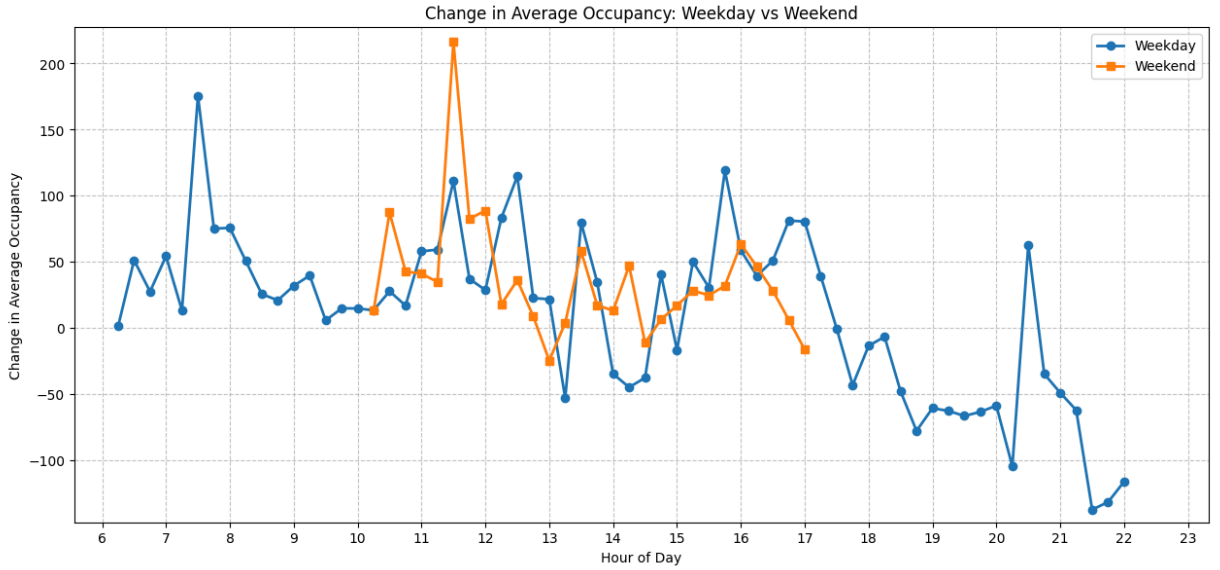

Figure 4: A graph of change in occupancy over time

4. Conclusion

4.1. Findings

To answer our original question, the best time to go to the rec is early Sunday morning. Remember, the times with the fewest people are in the early morning before the end of the 6-8

growth period and the days nearest the end of the week. For people with already established programs or early morning classes, a simple solution is shifting the start of your program forward a day to two to avoid the busiest days of the week.

4.2. Limitations

- Losing two weeks of data is a significant portion of the dataset and may have affected the results.
- Students are only required to scan an ID on the way in, not out. This makes the true occupancy unknown because they don't know exactly how many people have left.
- The occupancy on the website doesn't truly reflect how full the Rec center feels. For example, a fitness class may inflate the occupancy, but because classes are limited to studios it won't affect regular occupants.
- Fifteen minute increments lead to coarser data, which limited the quality of the data

4.3. Future Research

We hope to remedy some of the limitations of this study with future research. The first of which is checking how accurately the website tracks the number of people actually in the rec. This would be done with one person, sitting by the entrance counting people coming in and leaving. Then, comparing what they counted with what the website said at the same time. Collecting more data would also be helpful. For example, getting an entire semester of occupancy data would lead to better analysis.

Another avenue of research is needed into the other factors affecting when people can go to the rec. The most obvious would be tracking occupancy of dining halls along side the rec. The stagnations in usage growth occur around regular meal times, suggesting a significant number of people are eating. It's also important to consider how the distribution of class schedules affects when people can work out.

Finally, interviews with students may provide reasoning behind the usage patterns we have observed. The current research has a lack of qualitative data which would be solved by interviews or questionnaires. Interviews or questionnaires may also reach students who don't regularly use the rec center.