Exploring Attendance Patterns

Mahati Gorthy

CMSC320: Introduction to Data Science

May 12, 2025

1 Introduction

Class attendance has always been a persistent issue in lectures that don't mandate presence or participation. There are many factors that affect this, ranging from weather, personal problems, commuting status, lack of motivation, ease of coming to class, and more. In this report, I explore how weather affects class attendance at the University of Maryland, or more specifically in Professor Max Marowski's Spring 2025 CMSC 320 Introduction to Data Science lecture. By looking at data on both attendance and weather, I try to understand whether weather conditions—like rain or temperature changes—play a role in whether students show up to class. I use linear regression to analyze the data, hoping to uncover any patterns that might help predict attendance based on the weather forecast. Overall, this report aims to see if weather is a significant factor in attendance, which could in turn help professors and the university make smarter scheduling decisions.

The dataset consists of two types of data: images and counts from two different classes this semester, the Monday-Wednesday class and the Tuesday-Thursday class. For the images, I counted each person and summed it up. I gathered all this data in a CSV file, storing the class size, count, and date. From these responses, I explored patterns. Some interesting findings include attendance on rainy days was, on average, about 12 students lower than expected, while dry days showed slightly elevated attendance. However, this difference was not statistically significant. Additionally, temperature showed no meaningful impact on attendance, and contrary to common assumptions, attendance did not increase in the days leading up to midterm exams. These results suggest that while weather may visually appear to influence class participation, the statistical evidence remains inconclusive.

2 Background

Class attendance is influenced by a variety of things—motivation, course material, and yes, even the weather. We've all had those days when the weather is bad, and the thought of skipping class seems a little more tempting. Previous research suggests that things like rain or extreme temperatures can keep students home, while nicer weather might encourage them to attend. This study looks into whether weather really does have an impact on attendance at the University of Maryland, with a specific focus on temperature and precipitation.

I pulled weather data for College Park, Maryland, from the Meteostat API, which provides historical weather information. Then, I matched this weather data with attendance records for each day of the semester. The idea is to see if things like a rainy day or a sudden temperature drop correlate with fewer students showing up to class. If we can spot a clear connection, it could provide useful insights for scheduling and help the university understand what might be affecting class participation.

3 Methodology

3.1 Data Cleaning

The attendance data had some junk entries, like rows labeled "shitpic," which definitely wouldn't be useful for our analysis. So, I removed those. I also found that the date column was in a string format, so I converted it to a proper datetime object so I could analyze trends over time.

To make sure the data was in a usable order, I sorted the attendance by date and created a new column, day-number, to give each entry a sequential number. This way, I could track the overall trend in attendance throughout the semester. I also addressed any missing or strange values (like negative attendance counts), so the data would reflect the real picture as accurately as possible.

The weather data from Meteostat had a datetime index, so I had to reset it and turn it into a regular column. Then, I merged this cleaned-up weather data with the attendance data, matching dates so that I could analyze how each day's weather might be impacting class attendance.

4 Findings

4.1 Does Rain Affect Attendance?

To determine whether rain has an effect on attendance in CMSC 320, I first merged class attendance records with daily weather data using the Meteostat API, aligning each class session with its corresponding weather conditions. I then created a rain indicator variable equal to 1 when precipitation (prcp) was greater than 0, and 0 otherwise.

However, raw attendance alone is not enough to evaluate the effect of rain. Attendance in college classes tends to decline over time naturally, regardless of the weather, as mentioned by Max. To control for this, I fit a simple linear regression model with day-number as the predictor and attendance count as the response. This allowed me to calculate residuals, representing how much actual attendance deviated from the expected trend on a given day.

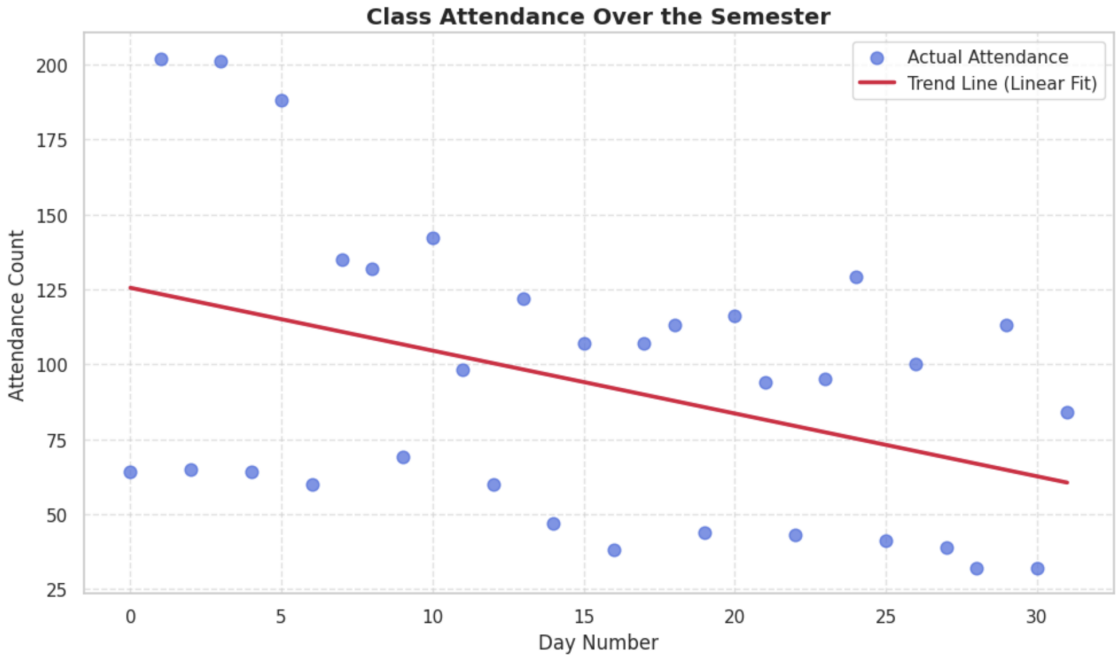

Figure 1: Attendance over time with trend line

Figure 1 shows actual attendance (scatter points) along with the predicted trend line from the regression model. The downward trend is clearly visible, reinforcing the need to model and remove this effect before testing weather-based questions and ideas.

After calculating the residuals, I compared their averages on rainy and dry days. On rainy days, the average residual was -12.21, indicating that attendance was about 12 students lower than expected. On dry days, the average residual was +7.33, suggesting a slight increase in attendance relative to the trend.

To visualize this, I plotted the residuals against time, colored by rain.

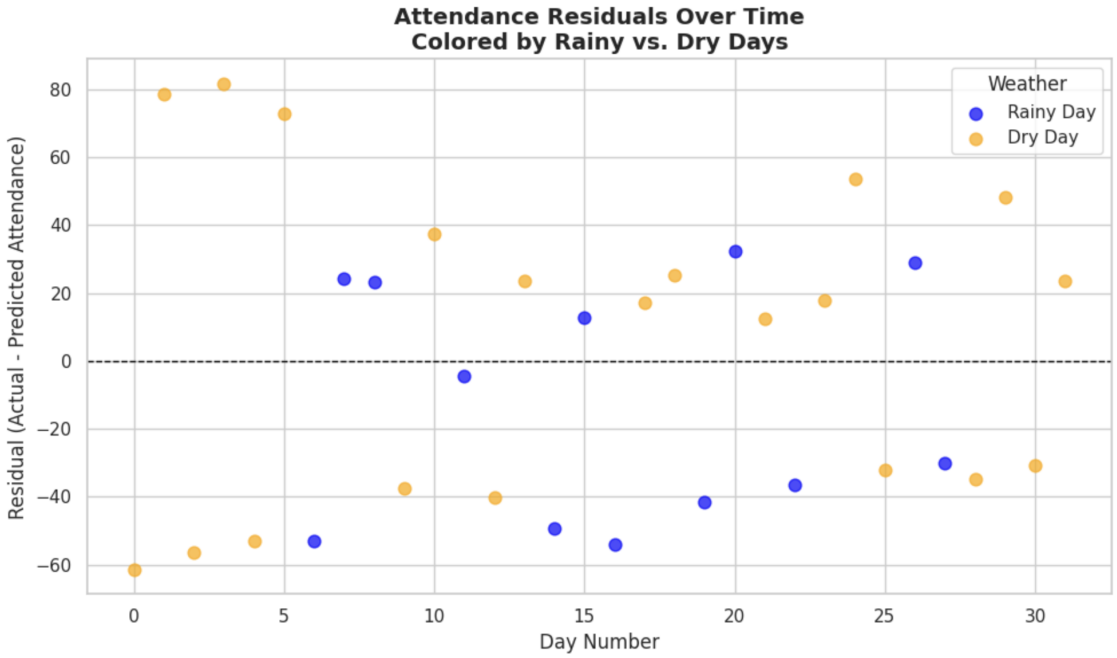

Figure 2: Residuals over time colored by rain (blue = rainy, orange = dry)

Figure 2 provides a clearer visual: blue dots (rainy days) frequently lie below the horizontal zero line, indicating a lower than expected attendance. Dry days (orange) fall more frequently above the line. This visual supports the hypothesis that rain may hurt class turnout.

However, to formally test this, I conducted a hypothesis test—a two-sample t-test comparing residuals on rainy vs. dry days. The test yielded a t-statistic of -1.34 and a p-value of 0.192. Since the p-value is greater than 0.05, we fail to reject the null hypothesis — the observed difference in attendance is not statistically significant. This means that the evidence is suggestive but not strong enough to confidently claim that rain affects attendance, in all cases.

4.2 Does Temperature Affect Attendance?

In addition to precipitation, I explored whether average temperature (tavg) influenced attendance. First, I plotted raw attendance counts against the average daily temperature.

Figure 3 shows no clear linear relationship between temperature and attendance. Although most days cluster between 0 and 20 ° C and show moderate attendance, there is substantial variation at all temperatures. In particular, there is no sharp drop-off in attendance on very cold or very hot days.

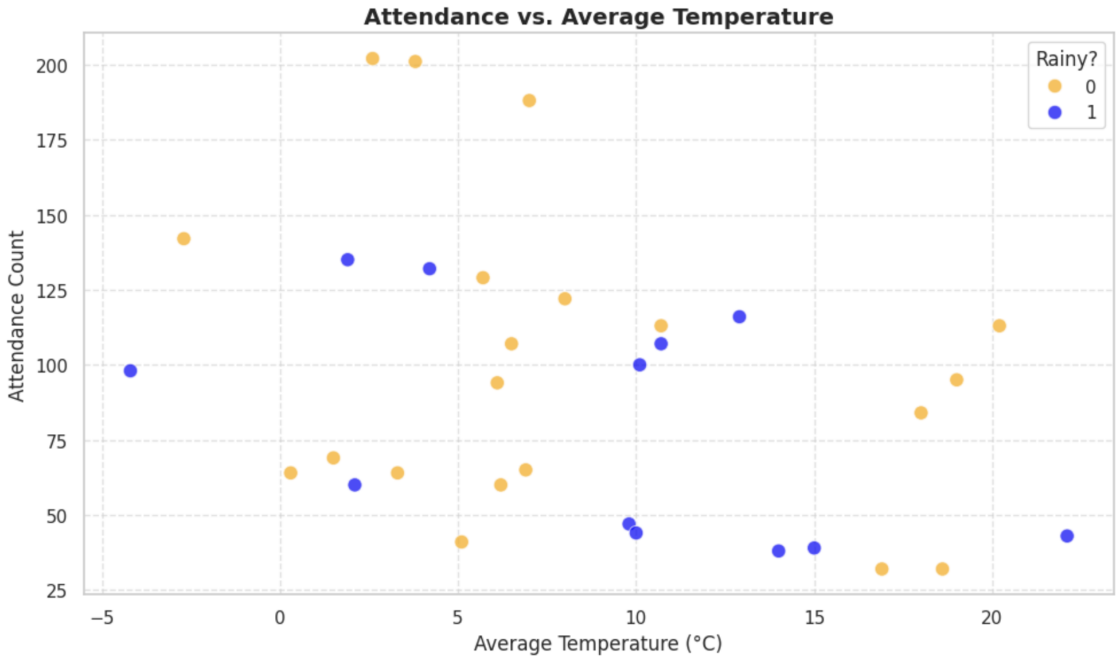

Figure 3: Attendance vs. average temperature (colored by rain)

Next, I divided class days into two categories: "High Temp" and "Low Temp", based on whether the temperature was above or below the median value. This let me compare attendance distributions across these groups using a boxplot.

Figure 4 shows that while the median attendance on "High Temp" days appears slightly higher, the interquartile ranges are wide and overlapping. There are also several outliers, indicating that temperature alone is not a strong driver of attendance. The boxplot supports the conclusion that there is no meaningful or consistent impact of temperature on attendance levels.

To further separate temperature's potential effect, I returned to residuals — values that remove the confounding effect. I plotted residuals against temperature, along with a regression line.

Figure 5 again shows no clear trend. The residuals fluctuate randomly at times, and the regression line is nearly flat, indicating a very weak (if any) relationship between average temperature and deviations from expected attendance.

Putting all of this together, these results suggest that temperature has no strong or consistent influence on attendance, whether considered directly or via residual analysis.

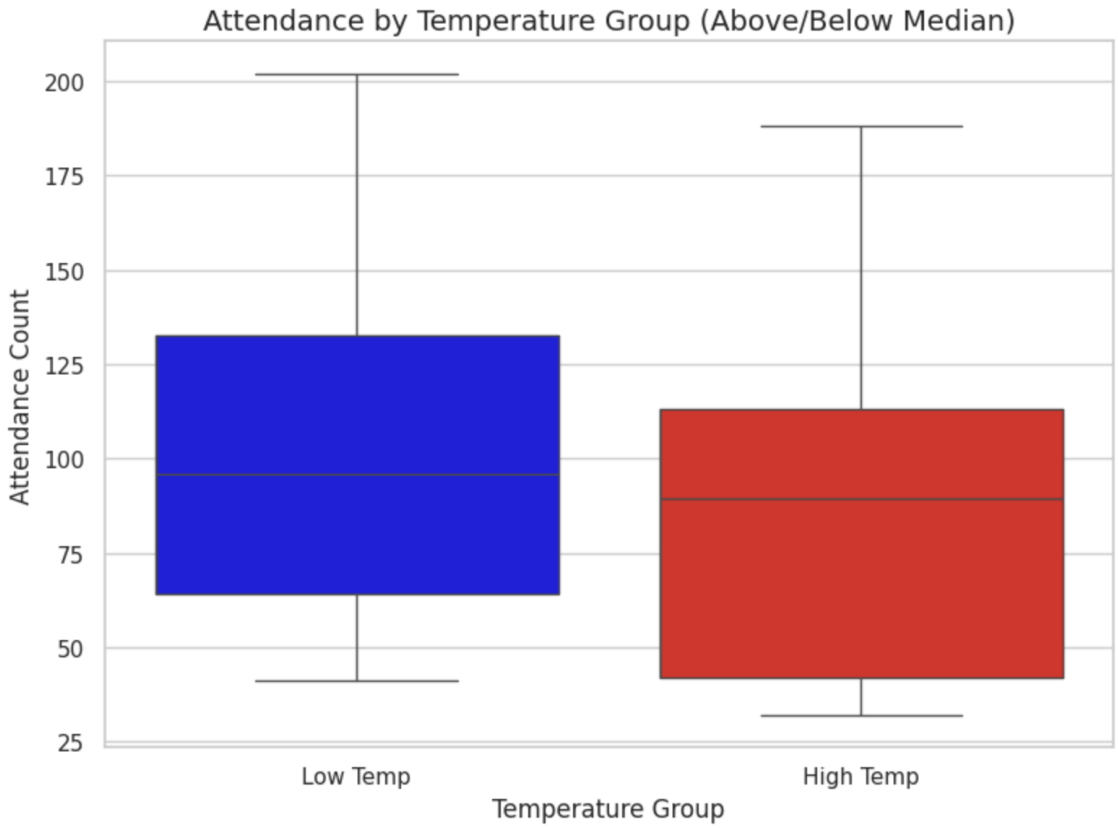

Figure 4: Boxplot of attendance by temperature group

4.3 Does Attendance Change Around Exam Classes?

I thought that this question would be interesting to ask— especially if certain exams were easier or harder. To investigate whether students attend class more often right before exams, I created a new column exam-period, tagging class days as "before-exam" if they occurred within three days before one of the two exam dates: March 12–13 (Exam 1) or April 16–17 (Exam 2). All other days were labeled as "neither".

Surprisingly, only 2 days fell into the "before-exam" category, and 30 days were labeled as "neither". I compared the average attendance for these two groups. "Before exam" days had a mean attendance of 76, whereas "neither" days had a higher mean of 94.13. This is the opposite of what one might expect — students showed up less before exams.

To visualize this, I created a boxplot comparing attendance counts for the two groups.

Figure 6 shows that the "before-exam" group has lower attendance and less spread, though the sample size is very small. To assess statistical significance, I again ran a t-test comparing the two groups. The resulting t-statistic was -0.477 with a p-value of 0.710, which is pretty far from significant.

Thus, based on this dataset, I found that no evidence that students attend class more

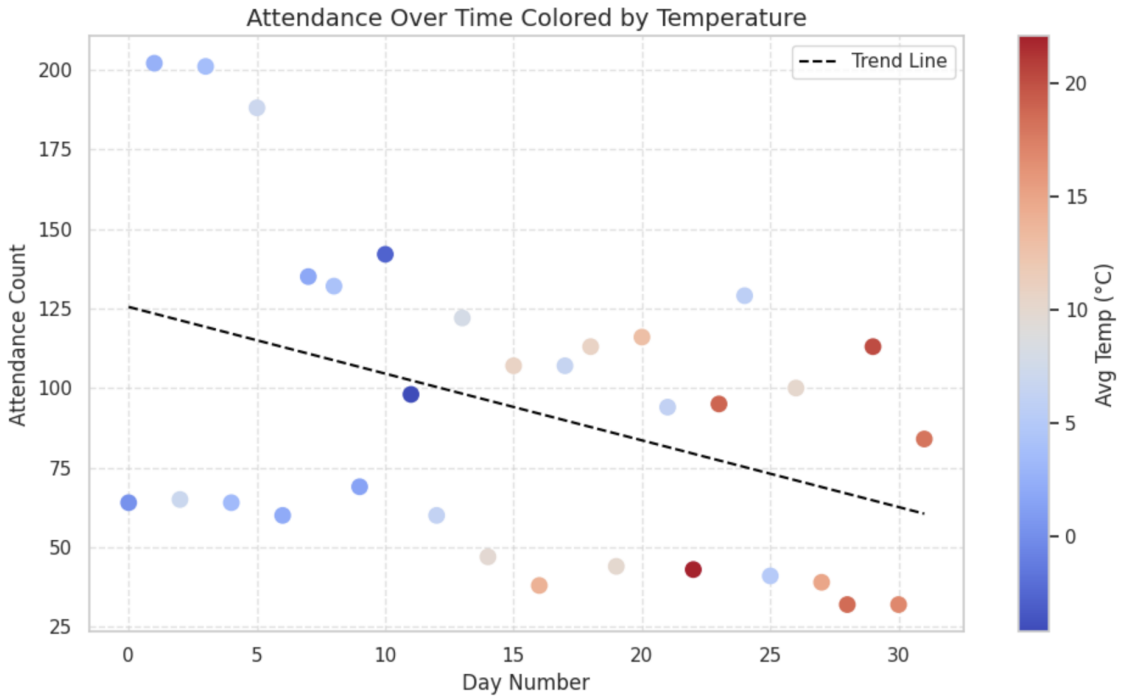

Figure 5: Residuals vs. temperature with regression line

in the days before exams. In fact, attendance may slightly drop — possibly due to students skipping lecture to study or attend other review/practice sessions with friends. However, with only two data points before exams, this analysis is likely not as supported.

4.4 Is there a correlation between attendance and other weather patterns?

To look at this question, I began by making sure that both the attendance and weather data shared a common datetime index. Once I had a single dataframe containing both, and all weather variables, I selected the numeric columns—including attendance, temperatures (tavg, tmin, tmax), precipitation (prcp), snow depth, wind direction (wdir), wind speed (wspd), and pressure (pres). I then generated a heatmap to visualize this.

Figure 7 shows the correlations between daily class attendance and various different weather variables. It is interesting to note that all three temperature values—average, minimum, and maximum—show moderate negative correlations (roughly –0.29 to –0.36) with attendance, implying that colder days tend to overlap with a higher turnout. Conversely, wind direction has the strongest positive relationship (+0.45), which may mean that certain wind patterns influence attendance. Precipitation had an interesting value (–0.15), while snowfall showed a slight positive correlation (+0.15). Other ideas such as wind speed and

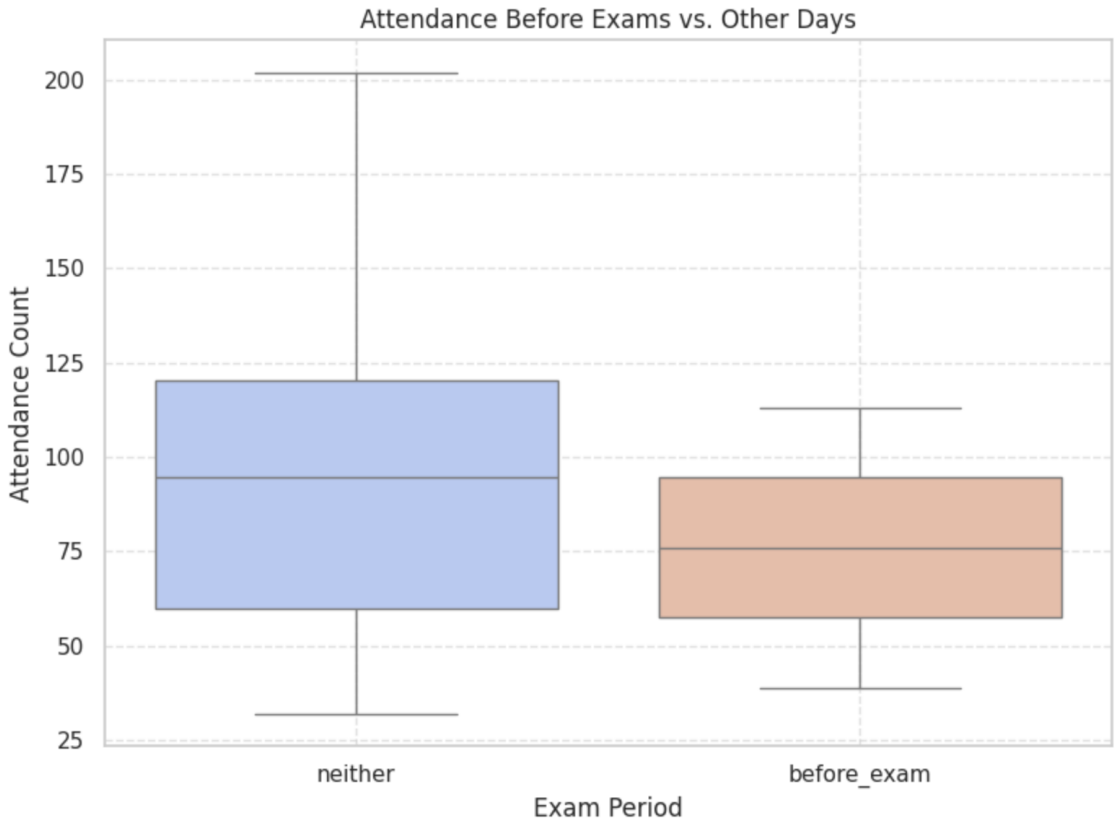

Figure 6: Boxplot of attendance before exams vs. other days

barometric pressure, have correlations near zero, meaning they have little impact on turnout. Finally, I didn't include the two variables ('wpgt' and 'tsun') because they contained only missing values. Overall, the heatmap highlights temperature and wind direction as the most promising predictors for a predictive attendance model, whereas the influence of other weather variables appears to be little to none.

5 Conclusion

To reiterate, this report set out to answer the question: Does weather affect class attendance in CMSC 320? To do so, I combined class attendance records with historical weather data from College Park and used statistical modeling and visualization techniques to isolate and evaluate the effects of rain, temperature, and exam timing on student attendance.

First, I controlled for the known natural decline in attendance over the course of a semester by fitting a linear regression model predicting attendance from time. This allowed me to calculate residuals — deviations from expected attendance — which could then be attributed to other factors like weather.

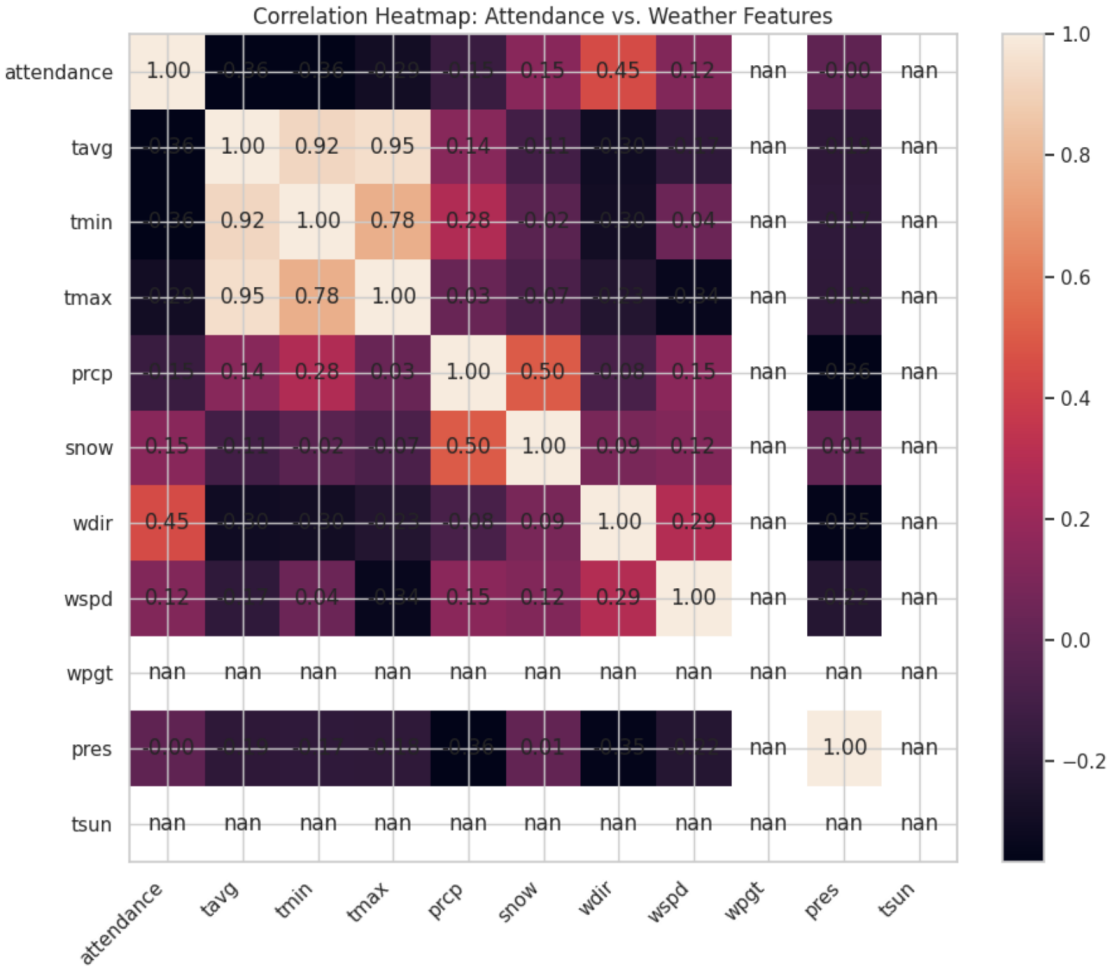

Figure 7: Heatmap of Attendance

Rainy days had noticeably lower residuals compared to dry days: students attended class roughly 12 fewer than expected when it rained. However, the difference was not statistically significant ($p = 0.192$), meaning we cannot confidently conclude that rain has a real effect on attendance based on this data alone. Temperature showed no meaningful correlation with attendance or residuals, and whether it was warmer or colder than usual had no consistent impact.

Interestingly, attendance did not increase in the days leading up to exams. In fact, average attendance was slightly lower before exams, and the statistical test showed no significance. This contradicts the common assumption that students attend more frequently before important assessments, though this result may be due to the small sample size of pre-exam days.

In summary, while there are suggestive visual patterns — especially related to rain — this semester's data provides no statistically significant evidence that weather or exam timing

affected CMSC 320 attendance. Further analysis with a larger dataset or over multiple semesters might reveal more conclusive trends.

6 Future Work

This analysis provides a foundation for understanding how environmental factors like weather may impact student behavior in the context of lecture attendance. However, there are several ways this work could be expanded or improved in the future.

First, the most significant limitation of this study was the sample size — only a single semester's worth of data was available, with very little coverage around key periods such as exams. Incorporating attendance data from multiple semesters and years would provide a much larger dataset and allow for more robust statistical conclusions. Seasonal variations in weather (e.g., fall vs. spring) and academic calendar differences could also be explored in terms of changes in weather.

Second, while precipitation and temperature were considered in this analysis, other weather variables such as wind speed, snowfall, or even perceived comfort (e.g., humidity index or wind chill) could provide additional insights—although I looked at some of these variables, a more in-depth analysis with a different API for more information can be interesting. It may also be worth examining more complex weather indicators, like storm alerts or drastic weather changes, which could influence student decisions more dramatically than mild rain.

Third, future work could integrate other contextual information beyond weather — for instance, the lecture topic of the day, whether a quiz or project deadline was approaching, or if alternative materials (e.g., lecture videos) were made available. These factors could be combined into a more robust model to better isolate the impact of weather alone.

Overall, expanding the dataset would significantly improve the explanatory power and precision of this analysis, opening the door to stronger conclusions and potential recommendations for course policy or scheduling.

7 Appendix

Dataset