**Weather and Campus Crime**

Jenna Anderson

**1. Introduction**

Campus safety is a constant concern for students, especially in public areas or at night. While discussions of crime typically focus on social, institutional, or demographic factors, some research suggests that environmental conditions—like weather—may also influence when and how crimes occur. Studies in environmental criminology and psychology have shown that temperature, humidity, and other atmospheric factors can impact human behavior, including aggression and decision-making[[1]](#footnote-1).

This project explores whether weather conditions—specifically temperature, humidity, and wind speed—are associated with the nature and timing of crimes reported on or near the University of Iowa campus. By merging university crime logs with localized hourly weather data, I wanted to see if any noticeable patterns would emerge—whether certain conditions tend to line up with types of crime, and whether weather could offer some small insight into when violent incidents are more likely to happen.

**2. Data**

This project uses two sources of information: crime reports from the University of Iowa and historical weather records from Cedar Rapids, Iowa. By combining these sources, I created a dataset that connects each reported crime with the weather conditions at the time it occurred.

*2.1 Campus Crime Log*

I scraped the crime log directly from the University of Iowa Office of Clery Compliance[[2]](#footnote-2). This log includes all reported criminal incidents that fall within the university’s Clery geography or within the patrol jurisdiction of the University of Iowa Police Department (UIPD). Each entry contains a case number, crime classification, the date and time the incident occurred, the date and time it was reported, location, and its case status.

I scraped the site using Selenium, initially collecting data from January through March 2025. However, since the university website only retains about three months of records at a time, I later revisited the site to capture the most recent entries and combined them with my original scrape to build a more complete and up-to-date dataset. Duplicate entries (based on case number) were removed.

*2.2 Weather Data*

For the weather data, I originally planned to use data from Iowa City’s own station, but unfortunately, it only retained about a month’s worth of history. Instead, I pulled three months of hourly weather records from Weather Underground[[3]](#footnote-3), using data from the Eastern Iowa Airport Station in Cedar Rapids. Given how close Cedar Rapids is to Iowa City (~25 miles), this seemed like a reasonable alternative.

The weather data included hourly observations of temperature, dew point, humidity, wind speed and direction, gusts, atmospheric pressure, precipitation, and general conditions (e.g., “Fair,” “Snow,” “Cloudy”). Like the crime data, I initially scraped three months of history and later appended a second scrape to cover new dates.

*2.3 Merging the Datasets*

One of the more complex parts of this project was the merging of the crime and weather datasets. Because crimes can occur at any time of day and weather data is reported at fixed hourly intervals, a direct join was not possible. Instead, each crime record was paired with the closest available weather observation on the same day by identifying the minimum time difference between timestamps.

After converting the crime and weather timestamps into consistent datetime formats, I created a new column in the crime dataset called “Nearest Weather Time,” which stored the matching hour-level time from the weather data. This ensured that even if a crime occurred slightly before or after an exact hour, it would still be aligned with the most relevant weather conditions.

Once this matching step was complete, I merged the datasets on both date and the nearest weather time. This produced a single dataset in which every crime incident was associated with temperature, humidity, wind speed, and other environmental factors recorded around the same time.

After the merge, I finalized the dataset by cleaning the weather fields (removing units like “°F” or “mph” and converting strings to floats) and dropped duplicate columns resulting from the join. I also created a new “Time Bucket” feature to categorize crimes by time of day—Morning, Afternoon, Evening, or Night—which would allow for clearer trend analysis in later sections.

Table 1 Data Dictionary

|  |  |  |  |
| --- | --- | --- | --- |
| **Column** | **Type** | **Source** | **Description** |
| Case Number | String | Crime Log | Unique identifier for each reported crime case. |
| Classification | String | Crime Log | Type of crime or violation reported (e.g., burglary, assault, liquor violation). |
| Date/Time Occurred | DateTime | Crime Log | The date and time the crime occurred. |
| Date/Time Reported | DateTime | Crime Log | The date and time the crime was reported. |
| Location | String | Crime Log | Location of the crime. |
| Disposition | String | Crime Log | Outcome or resolution of the case (e.g., closed, referred, inactive). |
| Date | Date | Both | Simplified date extracted from Date/Time Reported (YYYY-MM-DD format) (used for merging) |
| Time | Time | Both | Simplified time extracted from Date/Time Reported (HH:MM format) (used for merging) |
| Time Bucket | String | Both | Categorical grouping of the time of day (Morning, Afternoon, Evening Night) |
| Nearest Weather Time | Time | Both | The closest available weather observation time to the crime occurrence |
| Temperature | Float | Weather Data | Temperature (°F) at the time of the crime occurrence |
| Dew Point | Float | Weather Data | Dew point (°F) at the time of the crime occurrence |
| Humidity | Float | Weather Data | Humidity percentage at the time of the crime occurrence |
| Wind | String | Weather Data | Wind direction during the crime occurrence (e.g., WSW, SSE) |
| Wind Speed | Float | Weather Data | Wind speed (mph) at the time of the crime occurrence |
| Wind Gust | Float | Weather Data | Wind gust speed (mph) recorded near the time of the crime occurrence |
| Pressure | Float | Weather Data | Barometric pressure (inches of mercury) at the time of the crime occurrence |
| Precip. | Float | Weather Data | Amount of precipitation (inches) recorded at the nearest weather time |
| Condition | String | Weather Data | General weather condition description (e.g., Cloudy/Windy, Mostly Cloudy, Fair) |
| Violent Crime | Int (0 or 1) | Weather Data | Binary indicator: 1 = violent crime, 0 = non-violent crime (based on classification keywords) |

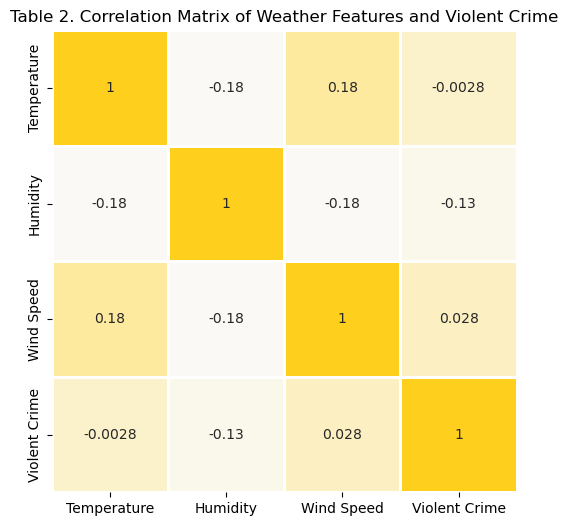
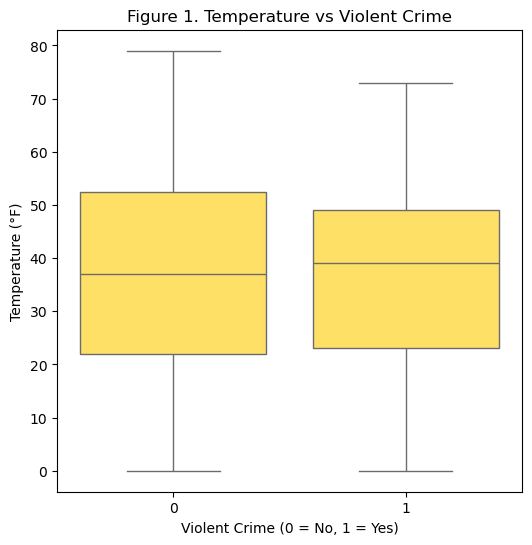
**3. Analysis**

*3.1 Temperature and Violent Crime*

One of the central questions of this project was whether temperature or other weather conditions are related to the likelihood that a crime is violent. Research has consistently suggested a link between heat and aggression; for example, a recent meta-analysis found that an 18°F increase in short-term temperature was associated with a 9% rise in violent crime incidents across international studies[[4]](#footnote-4).

To explore this locally, I created a binary indicator for violent crime and ran a **two-sample t-test** comparing temperatures during violent and non-violent incidents. The difference was not statistically significant (t = -0.06, p = 0.952), with nearly identical mean temperatures for both groups. This finding is further supported by the boxplot in **Figure 1,** which shows a substantial overlap in temperature distributions between the two groups.

To explore potential linear associations, I also calculated a **correlation matrix** across weather variables and the violent crime indicator. As shown in **Table 2,** the matrix revealed an almost nonexistent relationship between temperature and violent crime, with a correlation coefficient of approximately **-0.0028.** This further supports the earlier statistical finding that temperature alone is not meaningfully associated with whether a crime is violent.



*3.2 Time of Day and Violent Crime*

After finding no significant association between temperature and violent crime, I turned to time of day as a potentially more meaningful contextual factor. Each crime in the dataset was assigned to one of four time buckets—Morning, Afternoon, Evening, or Night—based on the time it was reported.

To test whether violent crime occurrence is associated with time of day, I conducted a **chi-square test of independence**. The results were statistically significant (X2 = 31.92, p < 0.001), indicating that the distribution of violent and non-violent crimes varies by time bucket. This suggests that time of day plays a role in the likelihood of a crime being violent.

A graph of a graph with numbers

AI-generated content may be incorrect.From the contingency table, we see that:

* **Afternoon** and **Evening** periods have higher counts of violent crime.
* **Nighttime** crimes are overwhelmingly non-violent, with only 8 of 110 incidents classified as violent.

A chart with yellow squares and white text

AI-generated content may be incorrect.**Figure 2** visualizes these proportions, showing that crimes occurring during the night are least likely to be violent, while afternoon and evening crimes are more likely to occur during periods of high social activity—particularly in the afternoon and evening—when interpersonal interactions are more frequent. In contrast, nighttime shows a significant decline in violent incidents, possibly due to reduced public activity and fewer opportunities for confrontation. This pattern aligns with routine activity theory, which suggests that crime is shaped by the timing of daily human behavior. As Cohen and Felson note, “the timing of work, schooling and leisure may be of central importance for explaining crime rates.”[[5]](#footnote-5)

To add more detail, **Figure 3** presents a heatmap showing the distribution of individual crime classifications across the four time buckets. Theft and simple assault cluster around the afternoon, when people are most active, while nighttime is largely characterized by alcohol-related offenses. These patterns suggest that daily routines and social behaviors shape not just the timing, but the nature of campus crime.

*3.3 Predictive Modeling of Violent Crime*

A yellow and white squares with black text

AI-generated content may be incorrect.To explore whether weather conditions could *predict* whether a crime would be classified as violent, I trained a series of supervised machine learning models using features from the weather dataset: temperature, humidity, and wind speed. The target variable was the binary violent crime indicator defined earlier in the analysis.

A pie chart with a yellow circle

AI-generated content may be incorrect.I began with a **logistic regression model**, which is commonly used for binary classification problems. After splitting the dataset into training and testing sets (70/30), the model achieved an accuracy of 68.4%. However, this is misleading. A closer look at the classification report reveals that the model almost entirely failed to identify violent crimes, achieving a recall of just 2% for the positive class (violent = 1). The pseudo-R-squared value was essentially zero (-0.0003), suggesting that the model performed worse than a simple baseline classifier that only considers class proportions.

The issue is especially evident in the confusion matrix shown in **Figure 4**, where the model correctly predicted 92 out of 93 non-violent crimes but misclassified nearly all violent ones. This reflects a critical limitation of standard classifiers in imbalanced datasets—they tend to default to the majority class. As shown in **Figure 5**, only about 28% of crimes in the dataset were classified as violent, which means a model can achieve high overall accuracy simply by predicting “non-violent” every time. Unfortunately, this means the model is ineffective for its actual purpose—identifying violent crimes. In this context, the minority class is arguably the more important one to detect, making the model’s high accuracy misleading and its practical value limited.

A yellow squares with black text

AI-generated content may be incorrect.To test a non-parametric alternative, I also trained a **K-Nearest Neighbors (KNN)** classifier using the same features. The KNN model produced a slightly lower overall accuracy of 67.6% but showed improved sensitivity to violent crimes. It correctly identified 9 of the 43 violent cases (recall = 21%), which is slightly higher compared to the logistic regression. However, as seen in **Figure 6**, this improvement came with a reduction in precision. The KNN model introduced more false positives and still failed to capture most violent crimes. This highlights the underlying issue: weather features alone cannot predict whether a crime will be violent or non-violent, regardless of algorithm choice.

A yellow squares with black text

AI-generated content may be incorrect.Given the class imbalance in the dataset, I next trained a **weighted logistic regression model,** which applies higher penalties for misclassifying the minority class (violent crimes). This encourages the model to treat both classes more evenly. The weighted model achieved a lower accuracy of 52.9%, but a much more balanced performance overall. As shown in **Figure 7**, the model correctly identified 21 violent crimes—ten times as many as the unweighted model—increasing recall to nearly 49%. While there are more false positives for non-violent crimes, the trade-off is arguably worthwhile if violent crimes are considered more critical to detect.

Despite a negative pseudo-R-squared value (-0.1181), the balanced logistic regression outperformed both previous models in identifying violent incidents. It demonstrates that class weighting can shift the model’s behavior meaningfully, even if the features themselves are limited in predictive strength. These results reinforce the idea that weather data alone may not be sufficient for modeling violent crime, but that algorithm design choices like class weighting can significantly affect a model’s usefulness in sensitive applications like public safety.

**4. Conclusion**

This project set out to explore whether weather conditions and time of day are associated with the likelihood or nature of reported crimes on and around the University of Iowa campus. By combining university crime logs with hourly weather data from Cedar Rapids, I investigated patterns in temperature, timing, and crime severity, applying both statistical analysis and machine learning models.

1. *Is temperature associated with whether a crime is violent?*

Not in this dataset. A two-sample t-test found no significant difference in average temperature between violent and non-violent crimes, and the correlation was near zero. This finding may be influenced by the limited timeframe of the data, which spanned only winter and early spring months and may have lacked sufficient temperature variation to detect any meaningful relationship.

1. *Does the likelihood of violent crime vary by time of day?*

Yes. The chi-square test of independence showed a significant association between violent crime and time of day. Violent crimes were more frequently reported during the afternoon and evening, while nighttime incidents were predominantly non-violent. These patterns suggest that temporal context, particularly periods of increased social activity, may play a role in crime severity.

1. *What types of crimes occur during different times of day?*

A heatmap of individual crime classifications showed that certain offenses cluster around specific times. Theft and assault were most commonly reported in the afternoon, whereas alcohol-related crimes such as public intoxication and liquor law violations were heavily concentrated at night. These trends add further context to the findings above, indicating that time of day influences not only crime severity but also the type of offense most likely to occur.

1. *Can weather predict whether a crime is violent?*

Not reliably. Logistic regression and KNN models trained on temperature, humidity, and wind speed struggled to identify violent crimes, largely due to class imbalance. While the weighted logistic regression model showed some improvement in recall, the overall predictive accuracy remained low. These results suggest that weather features alone are insufficient for modeling violent crime.

This analysis was constrained by its short timeframe, which only covered winter and early spring. Expanding the dataset to include summer months, additional years, or other locations could reveal stronger patterns, especially with more variation in temperature. Future work might also explore other influences on crime, such as location-specific factors, student population density, or the timing of campus events.

1. Anderson, C. A. (2001). “Heat and Violence.” *Current Directions in Psychological Science*, 10(1), 33-38. <http://www.jstor.org/stable/20182687> [↑](#footnote-ref-1)
2. <https://safety.uiowa.edu/crime-log> [↑](#footnote-ref-2)
3. <https://www.wunderground.com/history/daily/us/ia/iowa-city/KCID/date/2025-1-1> [↑](#footnote-ref-3)
4. Choi, H. M., Heo, S., Foo, D., Song, Y., Stewart, R., Son, J., & Bell, M. L. (2024). Temperature, crime, and violence: A systematic review and meta-analysis. *Environmental Health Perspectives,* 132(10), 1-12. <https://doi.org/10.1289/EHP14300> [↑](#footnote-ref-4)
5. Cohen, L. E., & Felson, M. (1979). *Social change and crime rate trends: A routine activity approach*. American Sociological Review, 44(4), 588-608. <https://doi.org/10.2307/2094589> [↑](#footnote-ref-5)