

Carson and Ainsley Final Project Results

Exploratory Data Analysis (EDA)

To explore the question “During what time of day are wildlife collisions most common?” We began by visualizing the distribution of wildlife strikes across different time_of_day categories. The bar chart of the cleaned FAA dataset showed that most strikes occurred during the day, followed by night, with dawn and dusk seeing the lowest number of incidents.

Recognizing the potential ambiguity of that variable, we conducted a second analysis using the time variable. After extracting the hour from each strike record, we plotted a histogram showing the count of wildlife strikes by hour. This showed a spike between 7 AM and 11 AM, with elevated rates continuing into the early evening. Very few strikes were recorded in the late night or early morning hours. This could potentially relate to the amount of flights flying during each time of day.

We also examined how this pattern varied seasonally by plotting time-of-day strike counts by month. While daytime consistently showed the most strikes, some months, for example October, showed slight change from this, where night strikes briefly surpassed daytime ones. This seasonal variability may reflect deeper patterns and cycles of timing in nature.

Hypothesis and Testing Approach

To formally assess whether wildlife strikes are distributed evenly across times of day, we conducted a Chi-square goodness-of-fit test using the time_of_day variable.

Hypotheses:

- Null hypothesis (H_0): Wildlife strikes are evenly distributed across the four time-of-day categories (dawn, day, dusk, night).
- Alternative hypothesis (H_1): Wildlife strikes are not evenly distributed across those categories.

In the first test, we assumed a uniform distribution, expecting equal strikes in each category. A second test applied custom expected values based on time-of-day duration providing a more realistic baseline for comparison.

Results:

- **Test 1 (Equal probabilities):** $X^2 = 159,472$, $df = 3$, $p\text{-value} < 2.2e-16$
- **Test 2 (Adjusted probabilities):** $X^2 = 29,549$, $df = 3$, $p\text{-value} < 2.2e-16$

In both cases, the p-values are well below 0.001, indicating statistically significant deviations from expected distributions.

Conclusion

Based on both visual and statistical analysis, we conclude that wildlife strikes are not evenly distributed throughout the day. Strikes are significantly more likely during daytime hours, a finding supported by both exploratory graphs and chi-squared tests with extremely small p-values ($p < 0.001$). This pattern exists even after accounting for variations in daylight hours. These results suggest that time-of-day is a critical factor in wildlife strike risk and should be considered in FAA safety policies, flight scheduling, and bird hazard mitigation efforts. Seasonal variations, such as the October spike in night strikes, highlight the need for adaptive risk strategies based on migration cycles and regional wildlife behavior.