HIT140 – Foundations of Data Science

**Group Project Report**

**HIT-140 Assignment-3**

**Bat and Rat Behavioural Analysis**

Group-4 Members: Aarjit Acharya, Aanchal Gautam, Yukrit Dangol

Date: October 15, 2025

GitHub Repository: https://github.com/yukritdangol/HIT-140-Assignment-3-Group-4

Table of Contents

[1. Introduction 3](#_Toc211448465)

[2. Methodology 3](#_Toc211448466)

[3. Results 4](#_Toc211448467)

[3.1 Descriptive statistics 4](#_Toc211448468)

[3.2 Inferential statistics 4](#_Toc211448469)

3.3 **Figures………………………………………………………………………………………………………………………….5**

[4. Discussion and Limitations 9](#_Toc211448470)

[5. Conclusion 9](#_Toc211448472)

[6. Individual Contributions 10](#_Toc211448473)

[7. Appendices 10](#_Toc211448474)

# 1. Introduction

This project analyses observational data to investigate behavioural interactions between bats and rats. So, to address the problem, we went with methodologies as; Two research questions are addressed: (A) whether the arrival or presence of rats affects bat risk-taking behaviour, and (B) how seasonal variation influences bat activity and any interaction with rat presence. The datasets include detailed timestamps, behavioural labels (risk, reward), and environmental covariates. Where we arranged the datasets into a clean format arranging the columns, by removing duplicates and also managing empty columns in the provided data sets.

# 2. Methodology

In Data processing, we went with: arranging raw CSV files were cleaned to standardise datetime fields, remove duplicates, and creating derived features such as 'rat\_present\_at\_landing' and 'seconds\_after\_rat\_arrival'. Two cleaned datasets were produced and used for analysis.  
  
whereas, for the Analytical approach, we went with descriptive statistics, and visualisations were used. The cleaned datasets were used to explore distributions and temporal patterns. Then, taking a statistical approach analyzing the data, Inferential methods consisted of a contingency test (Fisher's Exact test) to analyze whether the presence of the rat was associated with bat risk, and logistic regression models to estimate the association of the predictors (seconds\_after\_rat\_arrival, hours\_after\_sunset, season, reward) with the binary outcome 'risk'. An interaction model (season \* seconds\_after\_rat\_arrival) was also fitted to test whether the temporal effect differed between seasons.

# 3. Results

The results contain the findings that we did in the assignments, all the numeric as well as the figurative outputs are mentioned in brief summary.

## 3.1 Descriptive statistics

Dataset 1 (bat-level events) contains 907 observations and 12 columns after initial cleaning. Key descriptive summaries include:

- Number of observations: 907.  
- Mean hours after sunset: 5.531863 (mean hours\_after\_sunset).  
- Median hours after sunset: 5.622778 (median hours\_after\_sunset).

- Overall proportion of risky landings: 0.494 (448 / 906).

## 3.2 Inferential statistics

Summary of statistical tests and model outputs (generalised findings):

3.2.1 Fisher's Exact Test on contingency table (rat\_present\_at\_landing × risk):

- Contingency counts (rat absent): risk=1 -> 445, risk=0 -> 455; (rat present): risk=1 -> 3, risk=0 -> 3.

- Fisher's Exact result: p = 1.000; odds ratio ≈ 1.022. Interpretation: the simple contingency test does not show evidence of an association between rat presence at the instant of landing and bat risk behaviour in this dataset.

3.2.2 Logistic regression (risk ~ season + seconds\_after\_rat\_arrival + hours\_after\_sunset + reward):

- Reward: strong negative association with risk (reward reduces odds of risky landing). In the fitted model, reward showed a large negative coefficient (highly significant, p < 0.001), with an estimated odds ratio ~ 0.05, indicating that reward presence is associated with substantially lower odds of risk-taking.

- Season: gives positive but not statistically significant in the main model of the file (coef ≈ 0.305, p ≈ 0.197), which provides limited but clear evidence that season alone changes risk after accounting for other covariates, keeping in mind that the overall seasons of the world are not the same.

- Seconds after rat arrival and hours after sunset: small coefficients and non-significant effects in the full model (p > 0.25), indicating that temporal proximity to rat arrival and time-after-sunset had little explanatory power when included alongside other predictors.

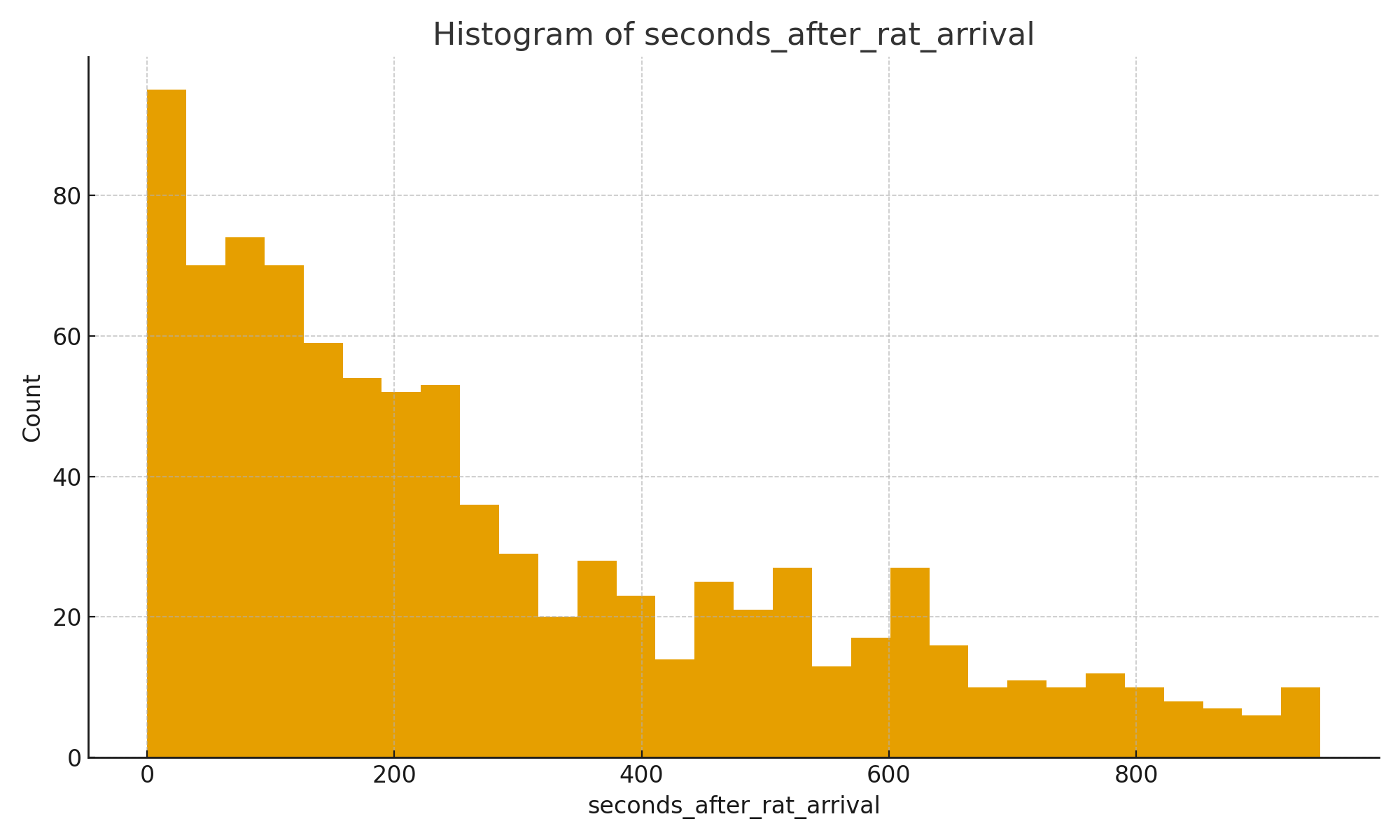
- Model fit: the logistic model used in the tests has a Pseudo R-squared ≈ 0.307 and a significant corresponding ratio test (LLR p-value < 0.001), indicating that the logistic model explains a meaningful and brief portion of variance compared to a model that intercepts only the reward variable, which is vastly conditional.

3.2.3 Interaction model (season × seconds\_after\_rat\_arrival):

- The interaction term between season and seconds\_after\_rat\_arrival was not found to be statistically significant (p ≈ 0.267) suggesting that there was no clear evidence that the effect of rat arrival on risk, as a function of time, differed by season. Reward also remained a strong predictor in the interaction model.

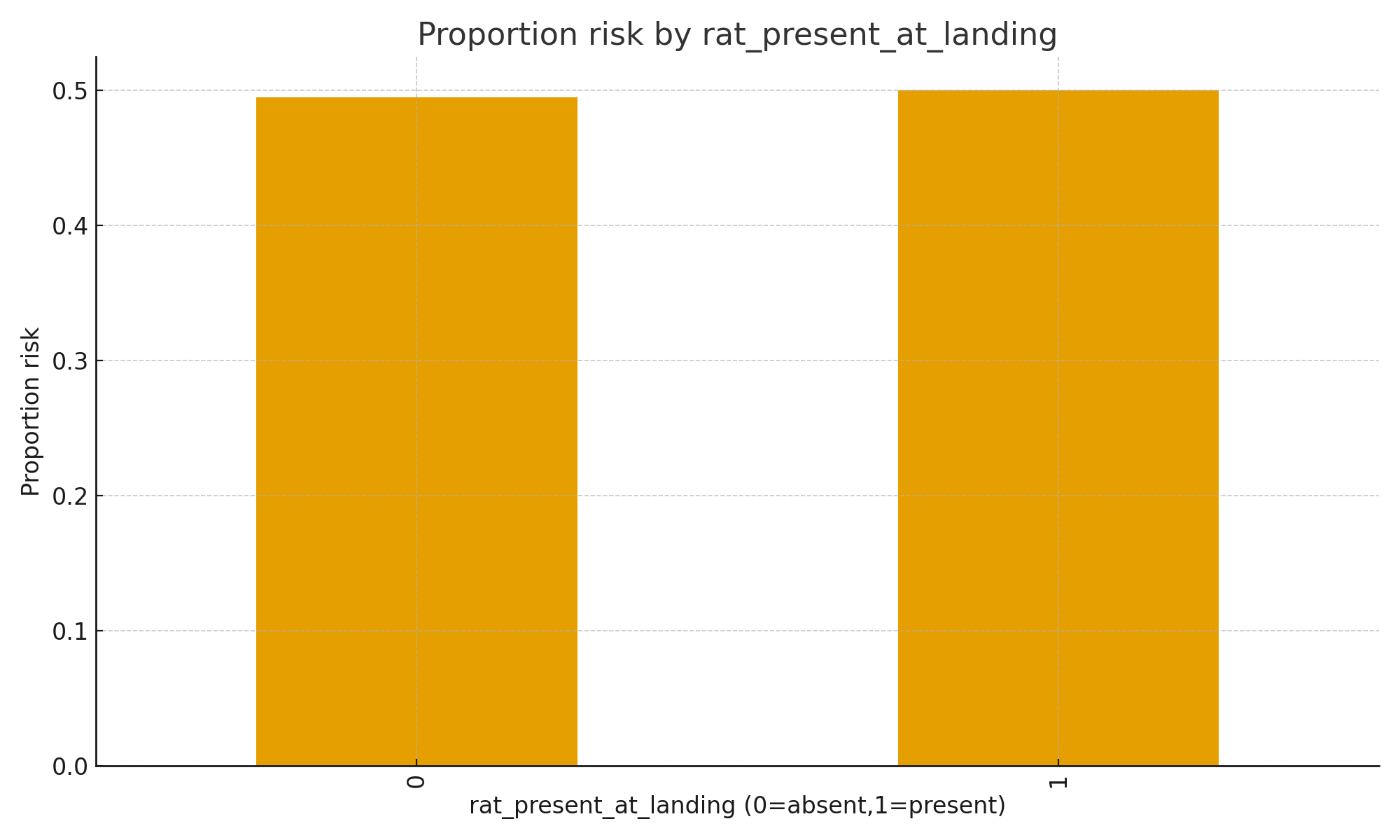
3.3 Figures

The figures generated from the statistical test for evidence are provided below;



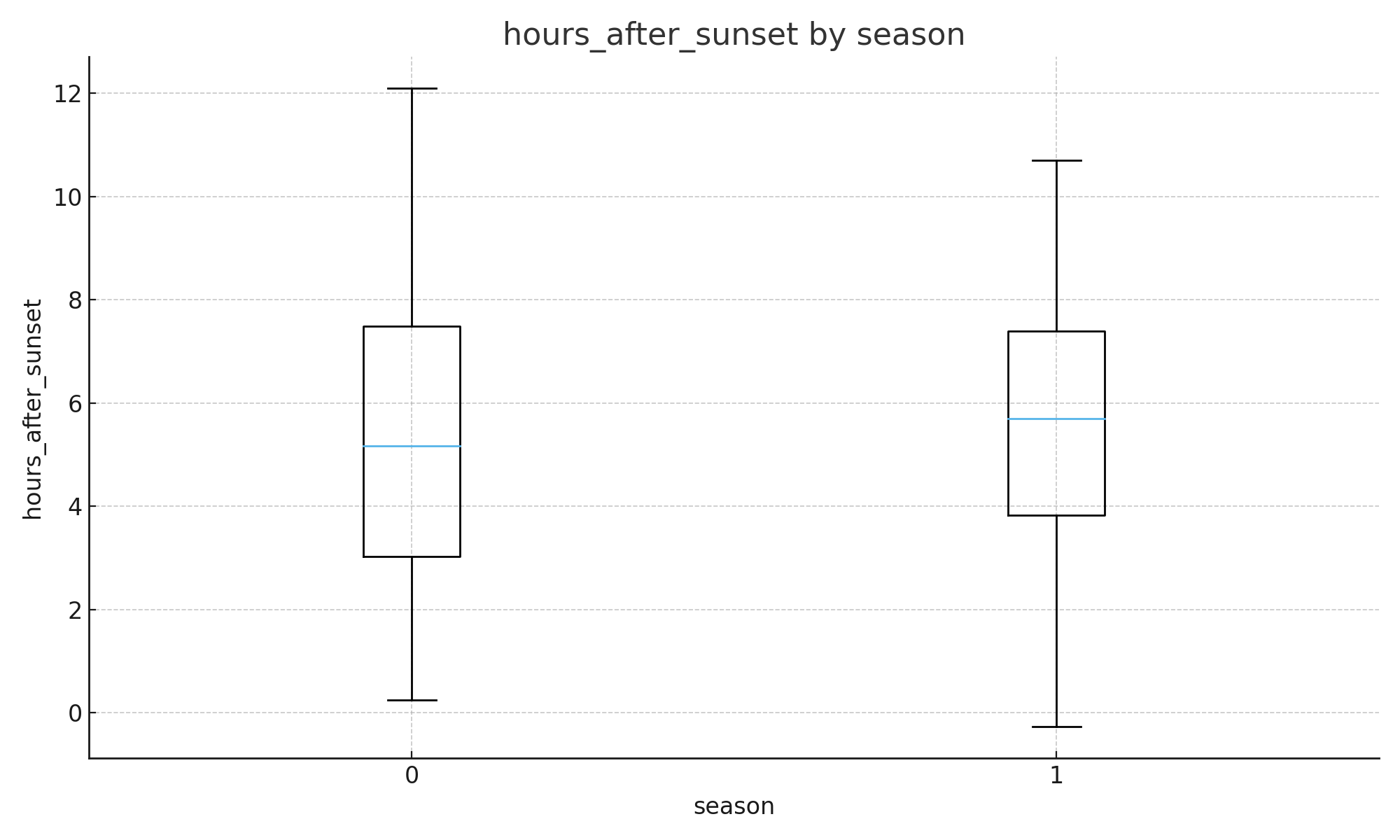
**Figure 1. Histogram of seconds after rat arrival**

Distribution of seconds between rat arrival and subsequent bat landings. The histogram shows a strong right-skew — many landings occur shortly after rat arrival, with a long tail of larger intervals.



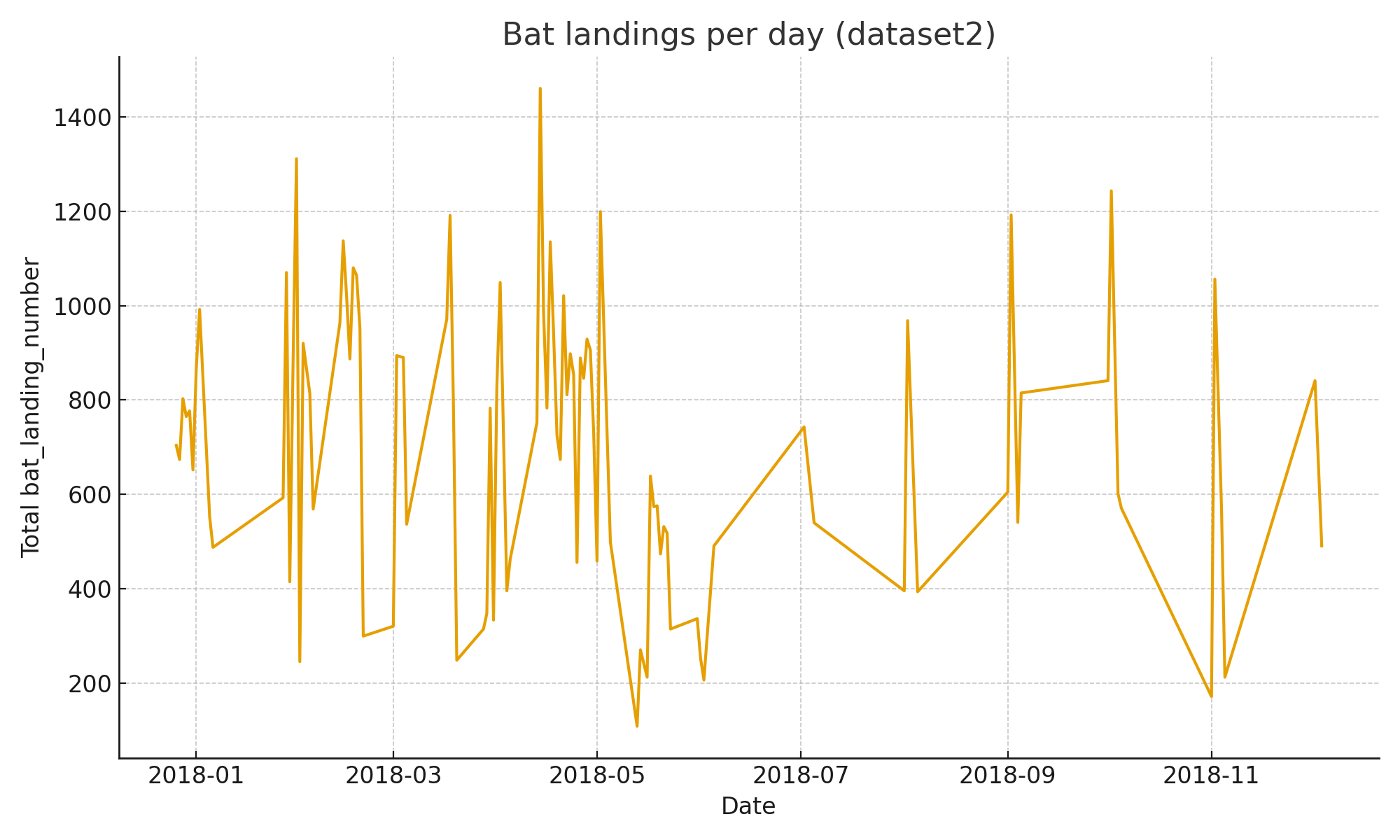
**Figure 2. Proportion of risky landings by rat presence**

Bar chart comparing the proportion of risk-taking landings when a rat was present versus absent. The visual shows near-equal proportions in this dataset (≈ 0.49–0.50), consistent with the Fisher test result of no association.



**Figure 3. Hours after sunset by season (boxplot)**

Boxplots of hours after sunset split by season categories (0,1). The median and interquartile ranges  
indicate modest seasonal differences in the timing of landings, but overlap suggests these  
differences are not strong.



**Figure 4. Time series of total bat landings per day (dataset2)**

A time series of bat landings in the observation period. The time series shows conditional peaks and valleys, which generally suggest that there may be variation in activity over time related to environmental or sampling effects.

# 4. Discussion and Limitations

# The analyses suggest that while reward presence predicted bat risk taking (rewarded landings were markedly less likely to be classified as "risk"), we had little evidence, in this dataset, of rat presence at the time of landing influencing risk probability. The Fisher exact test detected no relationship. Likewise, regression models that included temporal and seasonal covariates did not find significant effects of proximity to rats or time-after-sunset after controlling for reward. There are some important limitations to the conclusions: - Contingency counts of rat presence were imbalanced: there were very few landings observed at rat presence at the exact moment of landing thus limiting the power of contingency tests. - Any misclassification or timing of rat/bat timestamps may have confounded our ability to detect any temporal effects. - Observational data can not determine cause and effect — there may indicate some unmeasurable confounders to yielded pattern that we observed (i.e. local food availability, predator presence). - The reward variable is just so overriding in that it likely eclipsed any smaller ecological effects of rat presence.

# 5. Conclusion

This study took both descriptive and inferential approaches in exploring risk-taking in bats, influenced by the presence of rats and the context of seasonality. To summarize the findings (generalized): the presence of reward significantly decreased the odds of risk-taking, while the presence of rats at the landing and with respect to time showed no consistent significant effect after correction. Overall, the next steps should focus on increasing observations in the presence of rats, improving timestamps, and looking at things either causally or using longitudinal methods would help to disentangle the competing explanations more clearly.

# 6. Individual Contributions

Aarjit Acharya – Data Engineer / Cleaner:  
I managed the data cleaning, datetime parsing, deduplication, feature engineering (rat\_present\_at\_landing, seconds\_after\_rat\_arrival), and prepared reproducible scripts, for which I cleared duplicate data, managing columns with empty or duplicative data, and cleaned CSV outputs.  
  
Yukrit Dangol – Exploratory Analyst / Visualiser:  
Produced descriptive plots (histogram, bar chart, boxplot, time series), interpreted visual patterns, and prepared figure captions and notes for the report. Also code the main code to include every aspect of the research and also help manage the datasets and the findings of the team members.  
  
Aanchal Gautam – Statistician / Model Analyst:  
Conducted inferential analyses, including Fisher's Exact test and logistic regression (including interaction model), summarised statistical outputs, and drafted the Results and Discussion interpretation.

# 7. Appendices

Files and reproducibility:  
- Code and data: GitHub repository at https://github.com/yukritdangol/HIT-140-Assignment-3-Group-4 (file: HIT140\_group3\_submission\_with\_plots).  
- Analysis outputs: reports/appendix/analysis\_output.txt and reports/appendix/statistical\_tests.txt (key outputs summarised in Section 3).  
- To reproduce: see README\_RUN.txt in the repository (commands to run src/analysis.py and save figures to reports/figures/).