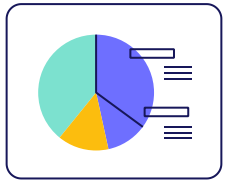


HIT140
FOUNDATIONS OF DATA SCIENCE

Assessment 2: Data Analysis

Darwin Group 23

September, 2025

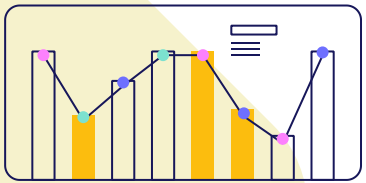
Team Members

Simin Li
S390115

Liang (Michael) Huang
S385350

Tianhui (Echo) Ke
S388810

Fuyu Li
S391488

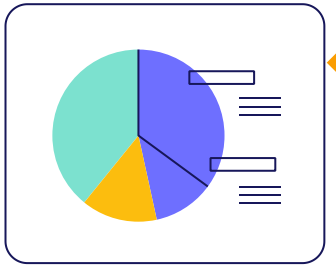


Content

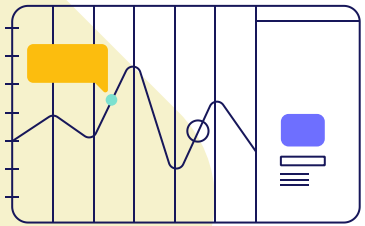
Background

A team of zoologists studied the foraging behaviour of the Egyptian Fruit Bats in the presence of Black Rats. They set out to conduct a series of observations in a semi_x0002_natural, open bat colony over a 7-month period, where the interactions of both animals' on a provisioned food platform were monitored using surveillance video cameras. The outcome of these observations was then collated into two datasets.

Objective



Do bats perceive rats not just as competitors for food but also as potential predators?



Hypothesis:

If rats are considered a predation risk by bats, scientists believe that this perception will translate into the bats' **higher level of avoidance behaviour** or **increased vigilance** during foraging on the food platform.

Data Overview

Dataset 1

Table 1. Variables in dataset1.csv

Column No.	Variable Name	Description
1	start_time	Bat's landing time on the food platform
2	bat_landing_to_food	Time difference (in seconds) after a bat's landing and before it started to approach the food on the platform
3	habit	The context surrounding the landing event and the animals' behaviour
4	rat_period_start	The time rat(s) arrived on the food platform
5	rat_period_end	The time rat(s) left the food platform
6	seconds_after_rat_arrival	Time difference (in seconds) since rats' arrival until the bat landed
7	risk	Indicates a risk-taking behaviour, such as attacking rats to gain access to food or not, as determined by the zoologist (0: risk-avoidance; 1: risk-taking)
8	reward	Whether the demonstrated behaviour is rewarding (0: no reward; 1: reward)
9	month	Month label assigned by the zoologists
10	sunset_time	Sunset time of the day
11	hours_after_sunset	Time difference (in hours) since sunset until the observed landing
12	season	Season label assigned by the zoologists

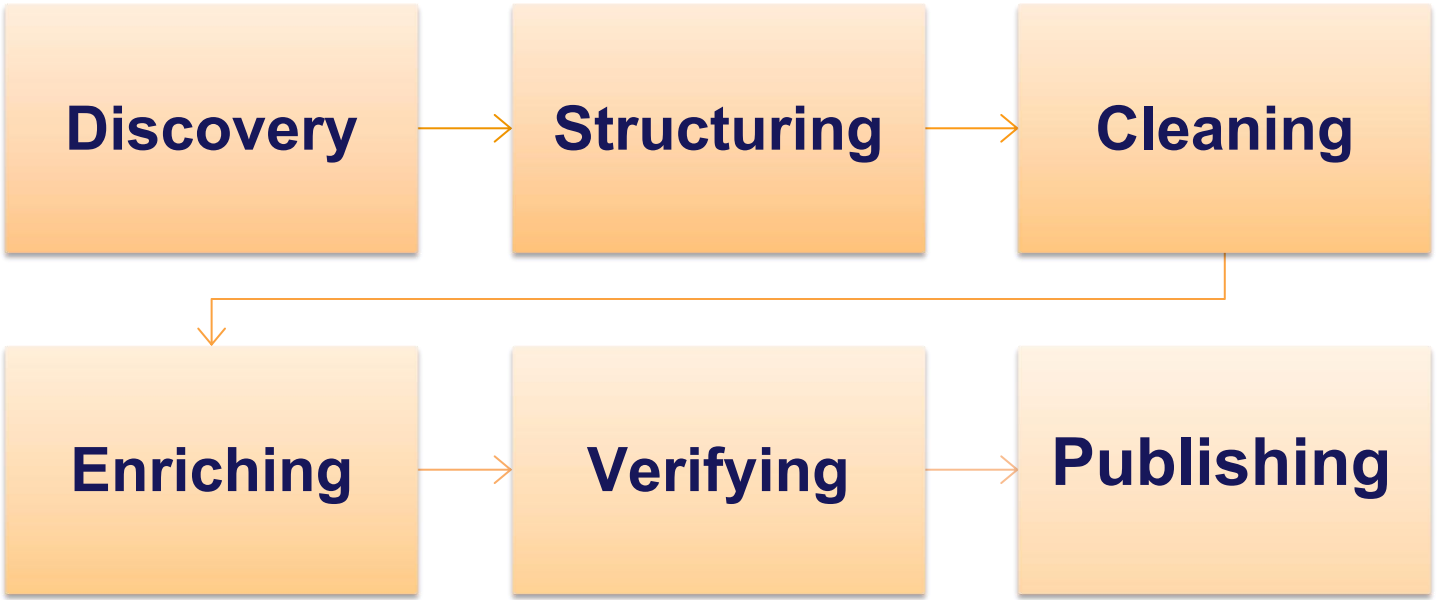
Data Overview

Dataset 2

Table 2. Variables in dataset2.csv

Column No.	Variable Name	Description
1	time	A 30-min observation period started
2	month	Month label assigned by the zoologists
3	hours_after_sunset	Time difference (in hours) since sunset until the observation period started
4	bat_landing_number	The number of bat landings counted within the observation period
5	food_availability	Amount of food remaining (unknown unit) that the zoologists estimated based on bat feeding activity.
6	rat_minutes	The aggregated duration of rat(s) were seen present on the food platform over the observation period
7	rat_arrival_number	The number of rat arrivals within the observation period

Data Wrangling Step



Data Wrangling for Dataset 1

No.	Method	Comment
1	Convert datetime columns	"start_time", "sunset_time", "rat_period_start", "rat_period_end" column
2	Convert numeric columns	"bat_landing_to_food", "seconds_after_rat_arrival", "risk", "reward", "hours_after_sunset" column
3	Drop rows with critical missing or invalid values	"start_time","risk", "bat_landing_to_food" column
4	Fill minor missing values	“hours_after_sunset” column
5	Check for duplicated rows	All the rows
6	Check numeric outliers	"bat_landing_to_food", "seconds_after_rat_arrival", "risk", "reward", "hours_after_sunset" (all numeric columns)
7	Time logic check	"rat_period_start", "rat_period_end" columns
8	Check and clean catagories column	“habit” column, create “habit_clean” column
9	Save cleaned dataset	dataset1_cleaned.csv

Data Wrangling for Dataset 2

No.	Method	Comment
1	Convert 'time' column to datetime	“time” column
2	Convert numeric columns	"month","hours_after_sunset", "bat_landing_number", "food_availability", "rat_minutes", "rat_arrival_number" column
3	Check basic info and missing values	No missing value
4	Save cleaned dataset	dataset2_cleaned.csv

Merging Dataset 1 and Dataset 2

No。	Method	Comment
1	Loading the dataset and parsing dates	"start_time", "sunset_time", "rat_period_start", "rat_period_end" column
2	Add “window_start” column in dataset1	“hour_after_sunset”, “sunset_time” in dataset1 → “window_start”
3	Align dataset2 column names	“time” → “window_start”
4	Merge datasets on window_start + month	"window_start", "month" column
5	Add log_latency	"bat_landing_to_food"
6	Rat pressure binning	“rat_minutes"
7	Add “rat_present_during_landing” column	"seconds_after_rat_arrival", "rat_minutes"
8	Clean habit column	Keep only bat behavior actions
9	Save merged dataset	merged_clean_dataset.csv

Descriptive Analysis

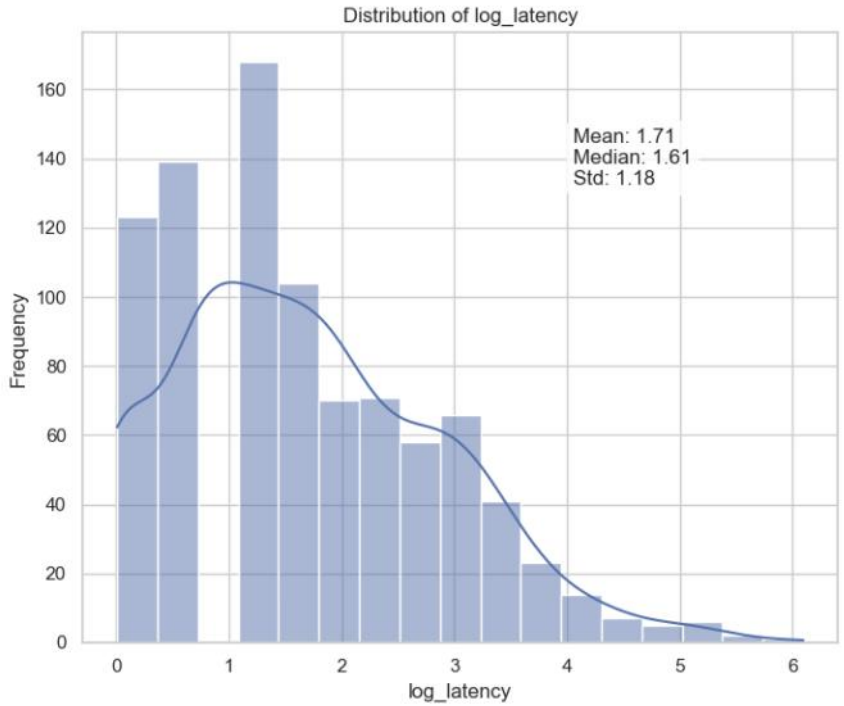
Methodology

- **Key variables analyzed:** log_latency, rat_minutes, rat_pressure_quartile, rat_present_during_landing
- **Cross-variable comparisons:** Explore relationships between rat presence/pressure and latency, risk-taking, and reward
- **Summary statistics:** mean, median, standard deviation, max, min
- **Visualization:** histogram, pie chart, bar chart, grouped bar plot, box plot

All analyses were conducted on the merged, cleaned dataset.

Descriptive Analysis

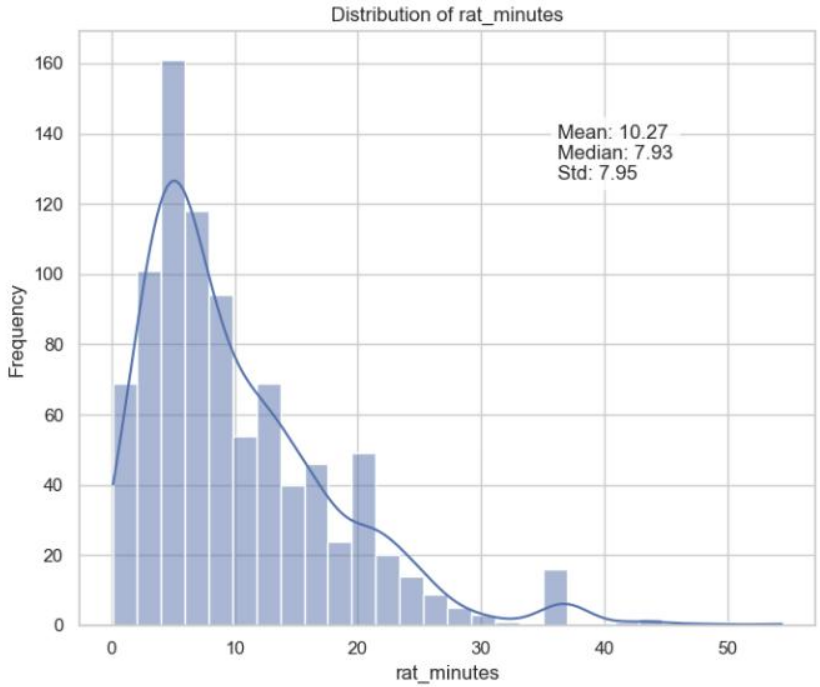
Key Variables Analysis: Log_latency



Taking the log of latency helps reduce magnitude and compress the scale without changing underlying relationships, making computation easier, stabilizing variance, and mitigating collinearity and heteroskedasticity in models.

Descriptive Analysis

Key Variables Analysis: rat_minutes, rat_pressure_quartile



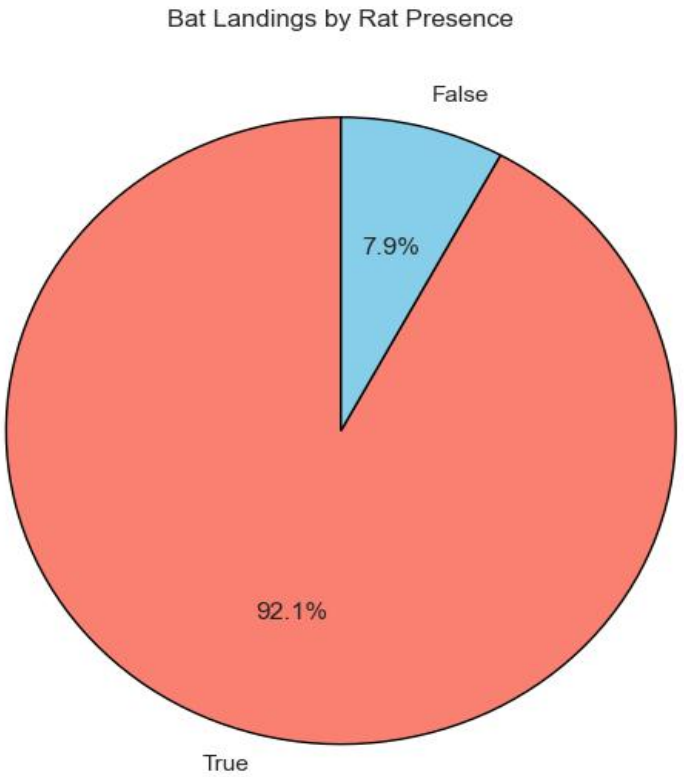
Rat Pressure Quartile Summary

rat_pressure_quartile	min	max	std	count	percentage(%)
Low	0.13	4.48	1.22	229.0	25.50
Medium	4.62	13.68	2.72	437.0	48.66
High	13.9	54.4	7.03	232.0	25.84
Total	-	-	-	898.0	100.00

Descriptive Analysis

Key Variables Analysis:
rat_present_during_landing

rat_present_during_landing	count	percentage
True	827.0	92.09
False	71.0	7.91
Total	898.0	100.00



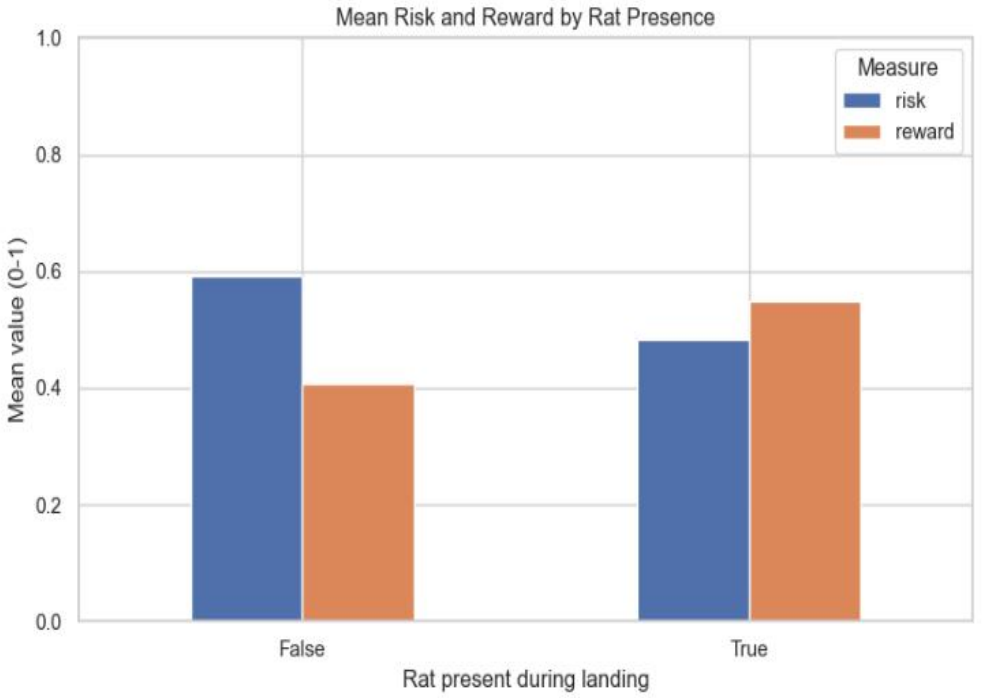
Descriptive Analysis

Risk vs Rat Presence

rat_present_during_landing	mean	std	count
False	0.591549	0.495046	71
True	0.483676	0.500036	827
Total	-	-	898

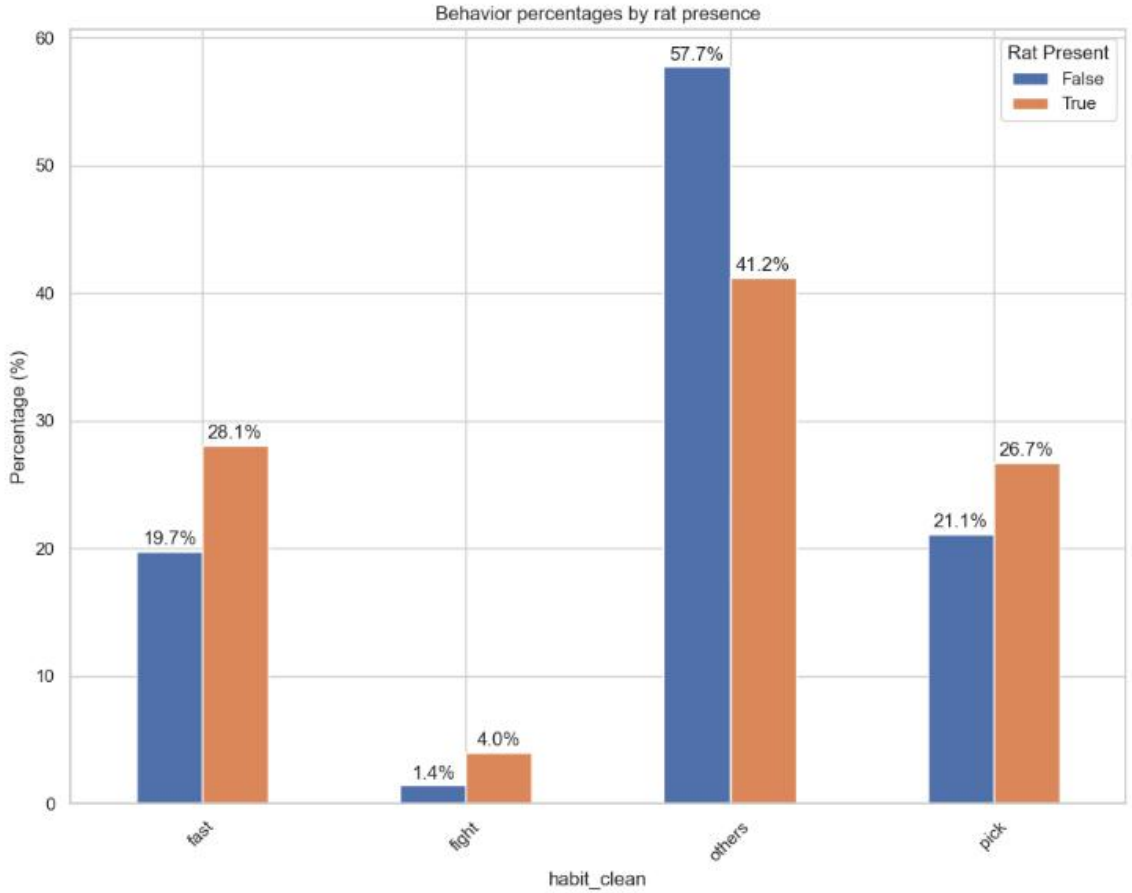
Reward vs Rat Presence

rat_present_during_landing	mean	std	count
False	0.408451	0.495046	71
True	0.547763	0.498015	827
Total	-	-	898



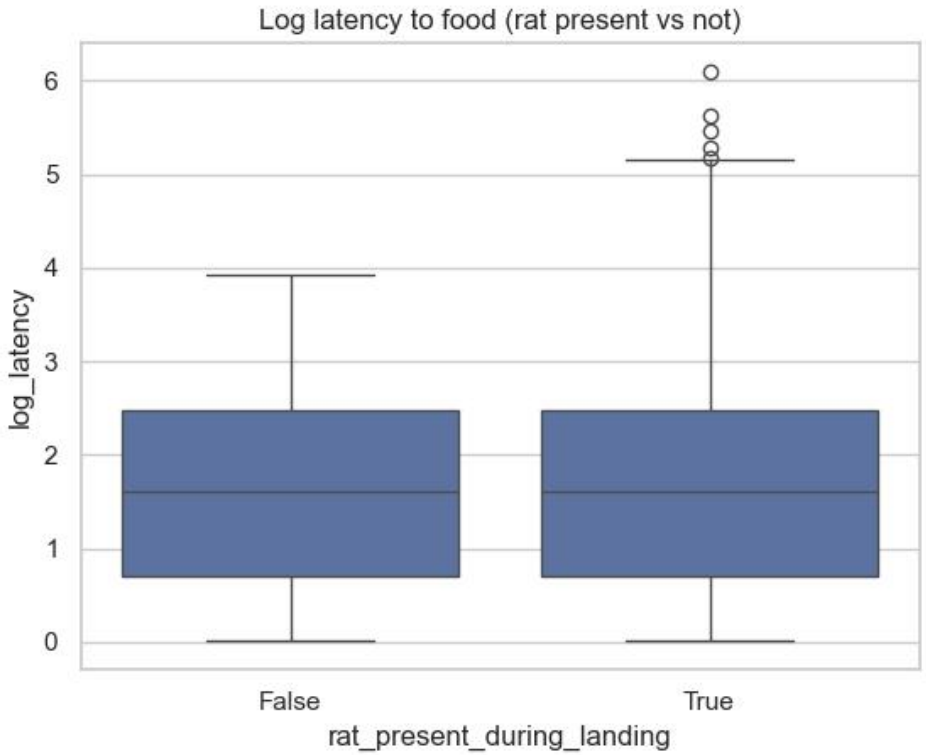
Descriptive Analysis

Risk vs Rat Presence



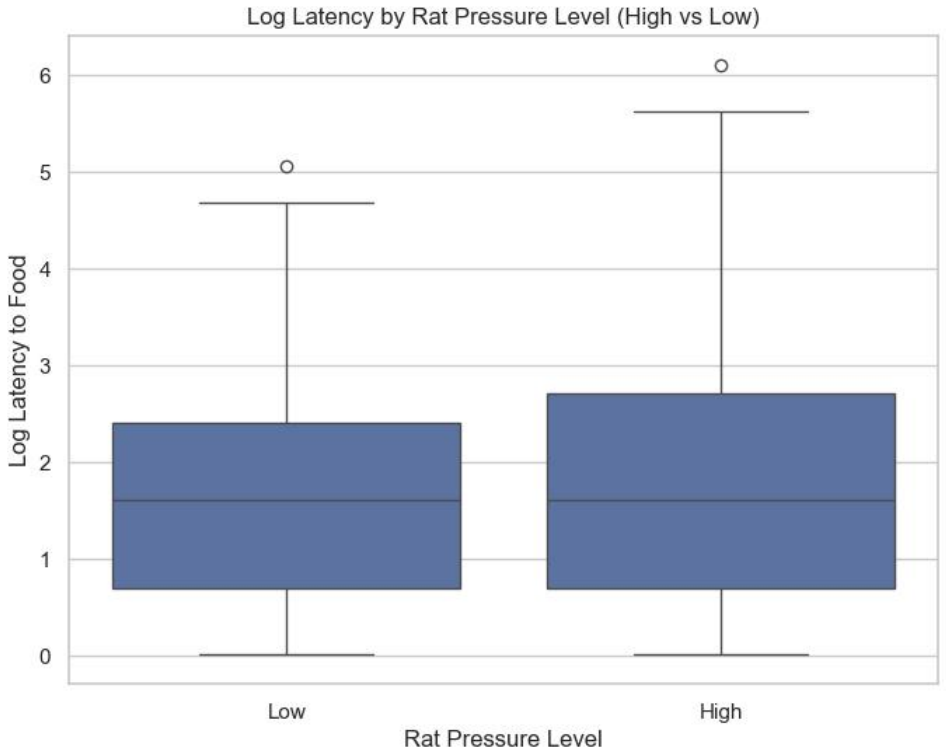
Descriptive Analysis

Latency vs Rat Presence



Descriptive Analysis

Latency vs Rat Pressure



Inferential Analysis

Do bats perceive rats not just as competitors for food but also as potential predators?

Methodology

- Hypothesis
- Z-score & Empirical Rule(To check normality)
- Mann-Whitney U test
- Chi-square test

Visualization:

- Histogram
- Boxplot
- Stacked Bar Chart

(All analyses were conducted on the merged, cleaned dataset.)

Inferential Analysis

1. Log Latency vs Rat Presence

Q: Do bats delay their landing more when rats are present?

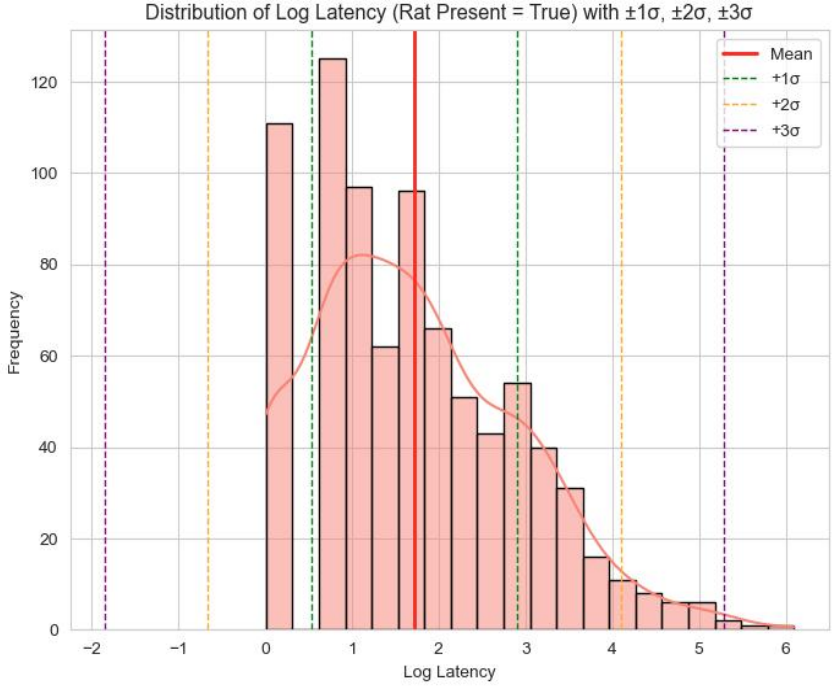
H0: Bat landing delay is not related to rat presence (True vs False).

H1: Bat landing delay is related to rat presence (True vs False).

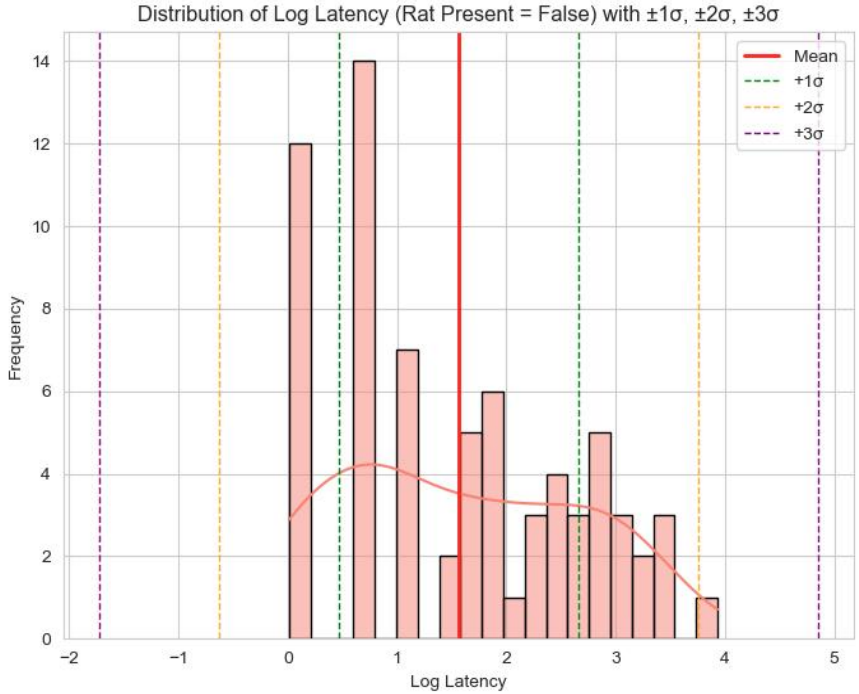
1. Latency vs Rat Presence

Normality Check (Using Z-score and Empirical Rule)

```
Normality Check (Z-score & Empirical Rule) for Log Latency (Rat Present = True)
Mean = 1.7194, Std = 1.1906
Within ±1σ: 68.92% (Expected ~68%)
Within ±2σ: 96.49% (Expected ~95%)
Within ±3σ: 99.64% (Expected ~99.7%)
```




```
Normality Check (Z-score & Empirical Rule) for Log Latency (Rat Present = False)
Mean = 1.5611, Std = 1.0954
Within ±1σ: 60.56% (Expected ~68%)
Within ±2σ: 98.59% (Expected ~95%)
Within ±3σ: 100.00% (Expected ~99.7%)
```



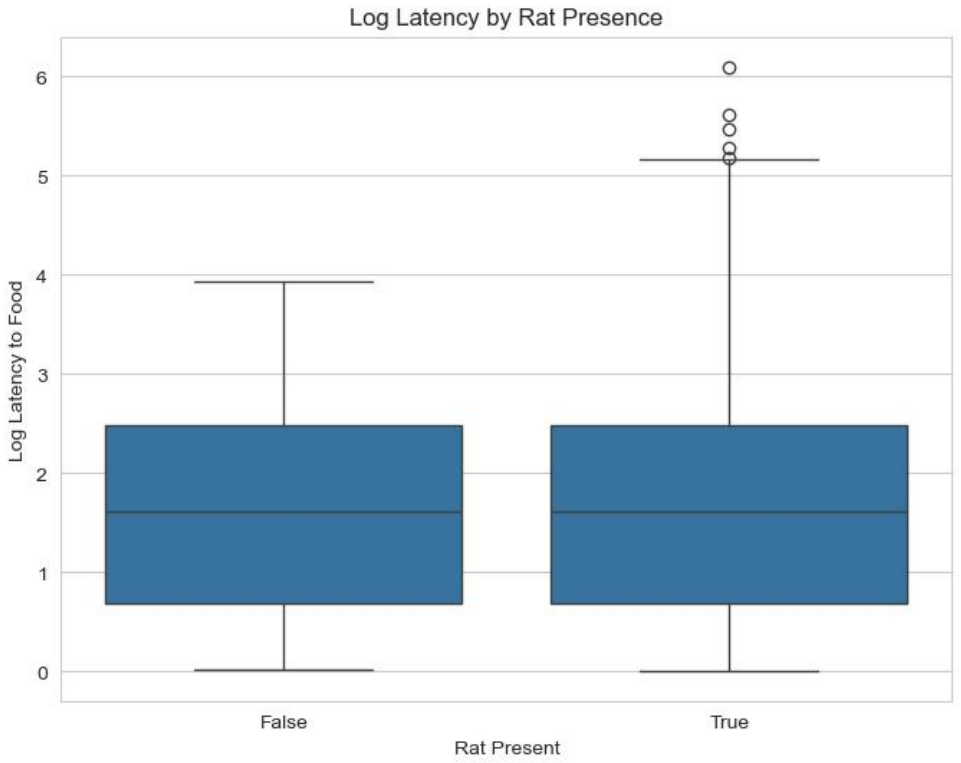
1. Latency vs Rat Presence

Mann-Whitney U test

U = 31205.50

p = 0.3772 > 0.05

Fail to reject H0. No evidence that bat landing delay is related to rat presence.



Inferential Analysis

2. Risk-taking Behaviour vs Rat Presence

Q: Are bats less likely to take risk when rats are present?

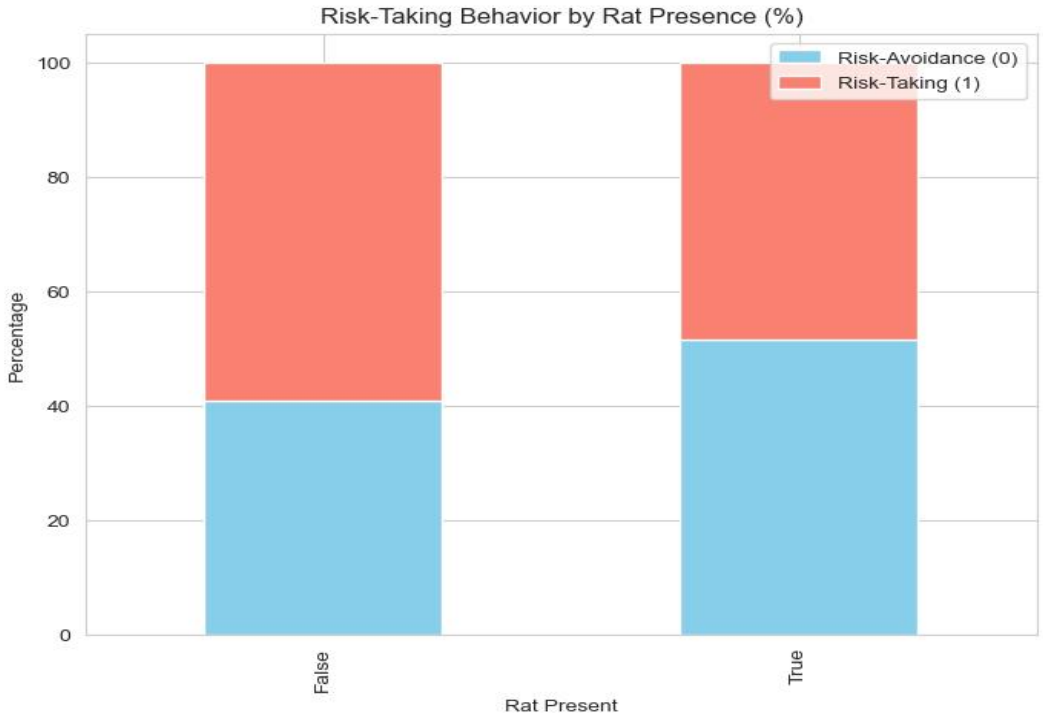
H0：Risk-taking behavior is independent of rat presence.

H1：Bats are more likely to avoid risks when rats are present.

2. Risk-taking Behaviour vs Rat Presence

Risk – Chi-square: $\chi^2=2.63$, $p=0.1050$, $df=1$ → **p >0.05**

Result: Fail to reject H0. No significant association between risk-taking behavior and rat presence.



Conclusion

Descriptive Analysis:

The presence of rats has little effect on bat latency but is associated with more cautious or rapid behaviors (“fast” and “pick”) and a marginal increase in risk-averse choices.

Inferential Analysis:

Mann-Whitney U and Chi-square tests indicate no statistically significant differences in latency or risk-taking behavior when rat present.

Overall Conclusion:

Only slight increases in vigilance and little change in latency, providing limited evidence that rats are perceived as potential predators.

Potential limitations in sample balance and data variability

Thank You!