Investigation A: Do Bats Perceive Rats as Predators?

This notebook addresses Objective 1 (Assessment 2) of the HIT140 Foundations of Data Science project.

We investigate whether bats perceive rats only as food competitors or also as potential predators.
If rats are considered a predation risk, bats should show more avoidance behaviour or increased vigilance during foraging.

We use two datasets:

```
1. dataset1.csv → Bat landings & behaviours (risk, reward, season, etc.)
2. dataset2.csv → Rat arrivals & bat activity (arrivals, food availability, etc.)
```

We will perform both:

```
1. Descriptive analysis (counts, trends, plots)
2. Inferential analysis (statistical tests: chi-square, correlation)
```

In [1]:

```
'''Importing libraries
    1. Pandas to process the provided data in dataset1 and dataset2
    2. matplotlib for plotting and visualization
    3. seaborn is a highlevel interface built on top of matplotlib which can work directly on pandas df
    4. scipy is used to perform scientific calculation such as correlation, chi square test etc.
'''
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.stats import chi2_contingency, pearsonr
```

In [2]:

```
# Configuring the specific style i.e graph background is white, muted colors and bigger font size
sns.set(style="whitegrid", palette="muted", font_scale=1.1)
```

In [3]:

```
# Loading the data sets using the pandas library
dataset1 = pd.read_csv("dataset1.csv")
dataset2 = pd.read_csv("dataset2.csv")
```

In [4]:

```
# Previewing the data sets for futhur cleaning and processing
print("Dataset1 (Bat Landings):")
# head function returns the first 5 rows of the data frame
display(dataset1.head())
# provides the detail information of the columns in the data frame
print(dataset1.info())
```

```
Dataset1 (Bat Landings):
```

	start_time	bat_landing_to_food	habit	rat_period_start	rat_period_end	seconds_after_rat_arrival	risk	reward	month	sunset_time
0	30/12/2017 18:37	16.000000	rat	30/12/2017 18:35	30/12/2017 18:38	108	1	0	0	30/12/2017 16:45
1	30/12/2017 19:51	0.074016	fast	30/12/2017 19:50	30/12/2017 19:55	17	0	1	0	30/12/2017 16:45
2	30/12/2017 19:51	4.000000	fast	30/12/2017 19:50	30/12/2017 19:55	41	0	1	0	30/12/2017 16:45
3	30/12/2017 19:52	10.000000	rat	30/12/2017 19:50	30/12/2017 19:55	111	1	0	0	30/12/2017 16:45
4	30/12/2017 19:54	15.000000	rat	30/12/2017 19:50	30/12/2017 19:55	194	1	0	0	30/12/2017 16:45

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 907 entries, 0 to 906
Data columns (total 12 columns):
 #   Column                    Non-Null Count  Dtype  
---  ------                    --------------  -----  
 0   start_time                907 non-null    object 
 1   bat_landing_to_food       907 non-null    float64
 2   habit                     866 non-null    object 
 3   rat_period_start          907 non-null    object 
 4   rat_period_end            907 non-null    object 
 5   seconds_after_rat_arrival 907 non-null    int64  
 6   risk                      907 non-null    int64  
 7   reward                    907 non-null    int64  
 8   month                     907 non-null    int64  
 9   sunset_time               907 non-null    object 
 10  hours_after_sunset        907 non-null    float64
 11  season                    907 non-null    int64  
dtypes: float64(2), int64(5), object(5)
memory usage: 85.2+ KB
```

In [5]:

```
# Previewing the data sets for futhur cleaning and processing same as dataset 1
print("\nDataset2 (Rat Arrivals):")
display(dataset2.head())
print(dataset2.info())
```

```
Dataset2 (Rat Arrivals):
```

	time	month	hours_after_sunset	bat_landing_number	food_availability	rat_minutes	rat_arrival_number
0	26/12/2017 16:13	0	-0.5	20	4.000000	0.0	0
1	26/12/2017 16:43	0	0.0	28	4.000000	0.0	0
2	26/12/2017 17:13	0	0.5	25	4.000000	0.0	0
3	26/12/2017 17:43	0	1.0	71	4.000000	0.0	0
4	26/12/2017 18:13	0	1.5	44	3.753857	0.0	0

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2123 entries, 0 to 2122
Data columns (total 7 columns):
 #   Column              Non-Null Count  Dtype  
---  ------              --------------  -----  
 0   time                2123 non-null   object 
 1   month               2123 non-null   int64  
 2   hours_after_sunset  2123 non-null   float64
 3   bat_landing_number  2123 non-null   int64  
 4   food_availability   2123 non-null   float64
 5   rat_minutes         2123 non-null   float64
 6   rat_arrival_number  2123 non-null   int64  
dtypes: float64(3), int64(3), object(1)
memory usage: 116.2+ KB
None
```

In [10]:

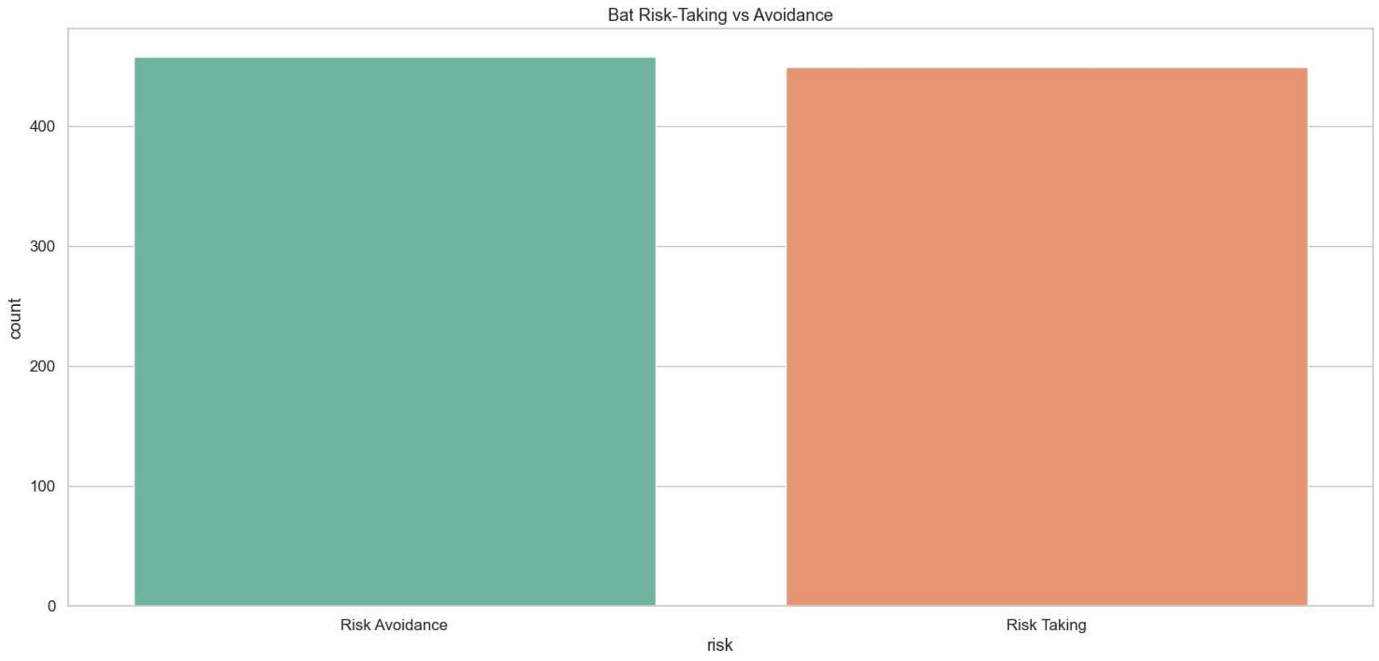
```
# Deteming the colums and their missing values for both data ets
print("Missing values in Dataset1:\n", dataset1.isnull().sum())
print("\nMissing values in Dataset2:\n", dataset2.isnull().sum())
```

```
Missing values in Dataset1:
 start_time                    0
bat_landing_to_food           0
habit                        41
rat_period_start              0
rat_period_end                0
seconds_after_rat_arrival     0
risk                          0
reward                        0
month                         0
sunset_time                   0
hours_after_sunset            0
season                        0
dtype: int64

Missing values in Dataset2:
 time                  0
month                 0
hours_after_sunset    0
bat_landing_number    0
food_availability     0
rat_minutes           0
rat_arrival_number    0
dtype: int64
```

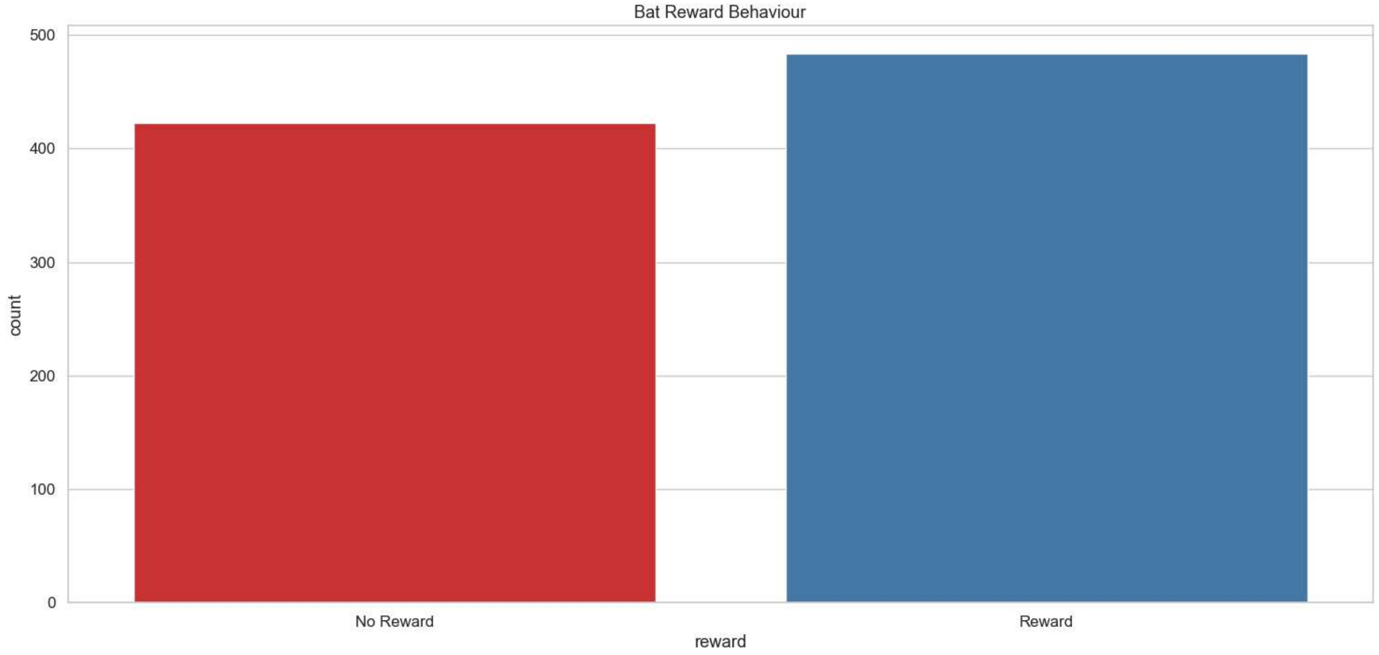
In [6]:

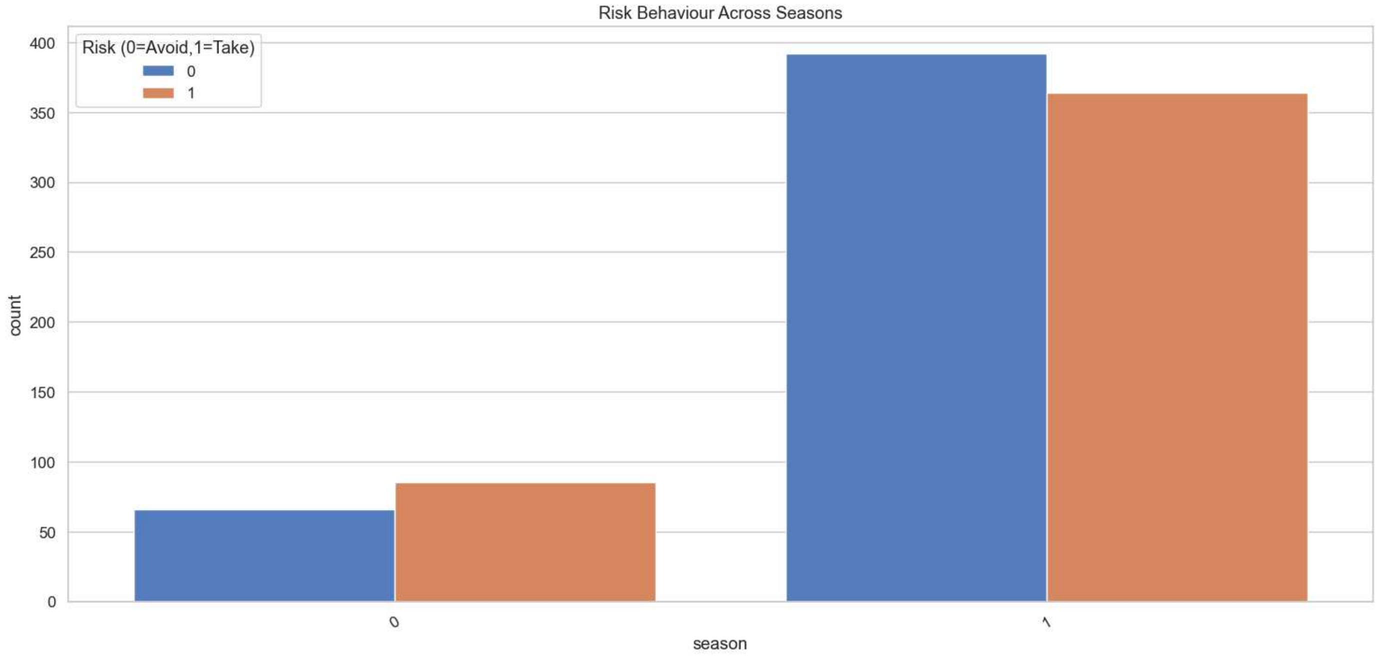
```python
# Descriptive Analysis: Dataset1 - Count risk-taking vs avoidance
# Renaming the boolen values in datasets i.e. 0 is represented as 'Risk Avoidance' and 1 is represneted as 'Risk
risk_counts = dataset1['risk'].value_counts().rename({0: 'Risk Avoidance', 1: 'Risk Taking'})
# Printing the values in the terminal
print("Risk Behaviour Counts:\n", risk_counts)

# Plotting the risk behaviour
# Defining the size of the graph
plt.figure(figsize=(18,8))
# Plotting a graph by looking into risk column for occurance of different values i.e. 0 and 1
sns.countplot(x='risk', data=dataset1, palette="Set2")
# Renaming the values for descriptive representation in the graph
plt.xticks([0,1], ['Risk Avoidance', 'Risk Taking'])
# Providing the heading name to the graph
plt.title("Bat Risk-Taking vs Avoidance")
plt.show()

# Renaming the boolen values in datasets i.e. 0 is represented as 'No Reward' and 1 is represneted as 'Reward'
reward_counts = dataset1['reward'].value_counts().rename({0: 'No Reward', 1: 'Reward'})
print("\nReward Behaviour Counts:\n", reward_counts)

# Plotting the reward behaviour and repeating the same process as for risk behaviour
plt.figure(figsize=(18,8))
sns.countplot(x='reward', data=dataset1, palette="Set1")
plt.xticks([0,1], ['No Reward', 'Reward'])
plt.title("Bat Reward Behaviour")
plt.show()

# Plotting the risk behaviour by season and repeating the same process as for risk behaviour
# Risk behaviour by season
plt.figure(figsize=(18,8))
sns.countplot(x='season', hue='risk', data=dataset1, palette="muted")
plt.xticks(rotation=30)
plt.legend(title="Risk (0=Avoid,1=Take)")
plt.title("Risk Behaviour Across Seasons")
plt.show()
```

```
Risk Behaviour Counts:
 Risk Avoidance    458
Risk Taking       449
Name: risk, dtype: int64
```



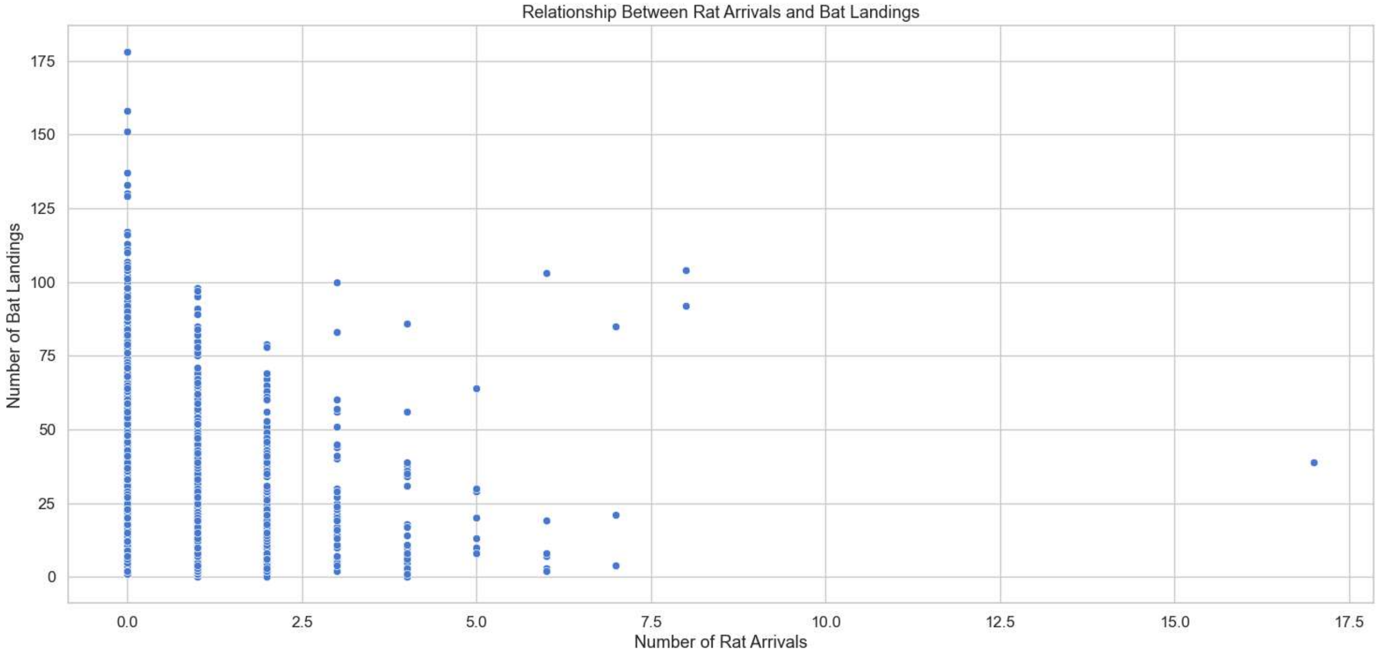
```
Reward Behaviour Counts:
 Reward       484
No Reward    423
Name: reward, dtype: int64
```

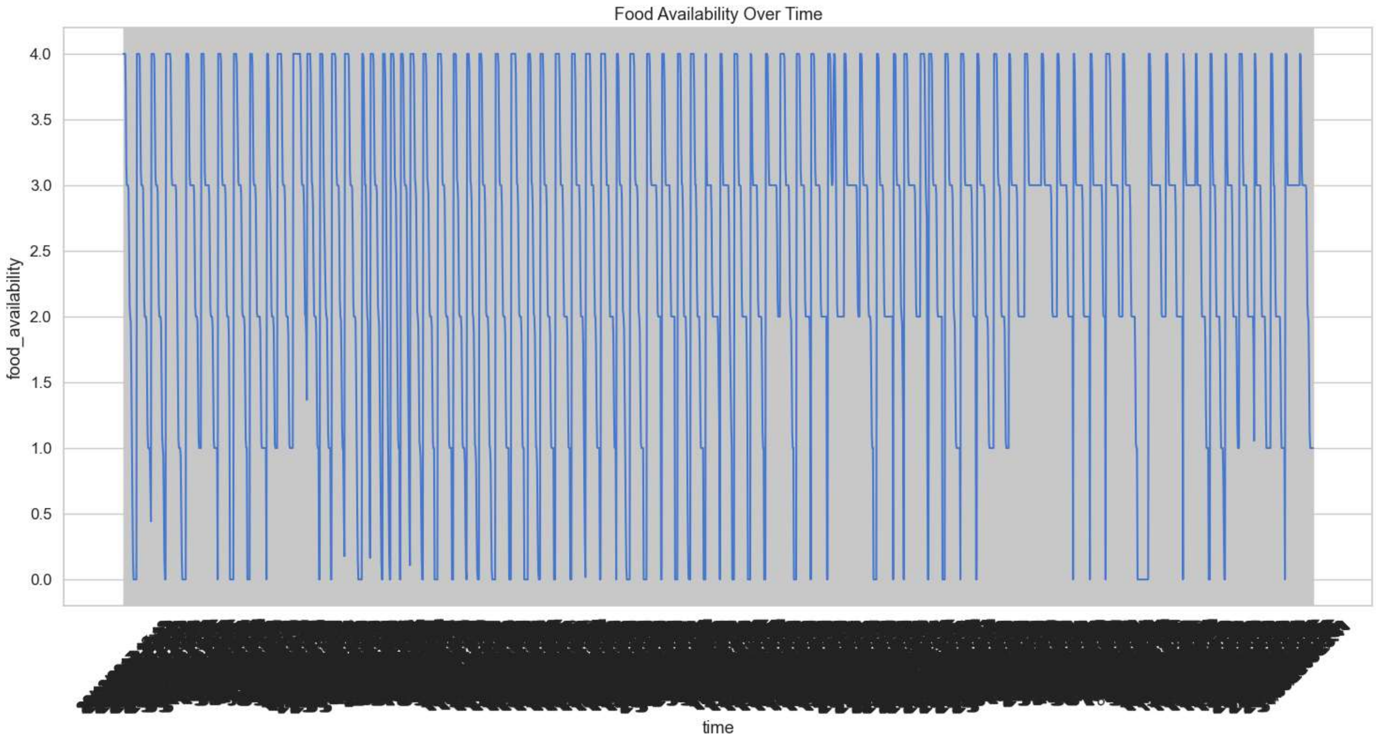




In [7]:

```
# Descriptive Analysis: Dataset2 - Scatter plot: Rat arrivals vs Bat landings
plt.figure(figsize=(18,8))
sns.scatterplot(x='rat_arrival_number', y='bat_landing_number', data=dataset2)
plt.title("Relationship Between Rat Arrivals and Bat Landings")
plt.xlabel("Number of Rat Arrivals")
plt.ylabel("Number of Bat Landings")
plt.show()

# Line plot: Food availability over time
plt.figure(figsize=(18,8))
sns.lineplot(x='time', y='food_availability', data=dataset2)
plt.xticks(rotation=45)
plt.title("Food Availability Over Time")
plt.show()
```





In [10]:

```
# Inferential Analysis
# 1. Chi-Square Test: Risk vs Reward

# Create contingency table
contingency_table = pd.crosstab(dataset1['risk'], dataset1['reward'])
print("Contingency Table (Risk vs Reward):\n")
print(contingency_table, "\n")

# Perform chi-square test
chi2, p, dof, expected = chi2_contingency(contingency_table)
```

```python
print("Chi-square Test Results (Risk vs Reward):")
print("Chi2 statistic:", chi2)
print("Degrees of freedom:", dof)
print("Expected frequencies:\n", expected)
print("p-value:", p)

if p < 0.05:
    print("\n✅ Significant relationship between Risk and Reward (p < 0.05)")
else:
    print("\n❌ No significant relationship between Risk and Reward (p >= 0.05)")


# 2. Correlation Test: Rat Arrivals vs Bat Landings
print("\nCorrelation Test (Rat arrivals vs Bat landings):")

corr, pval = pearsonr(dataset2['rat_arrival_number'], dataset2['bat_landing_number'])

print("Correlation coefficient (r):", corr)
print("p-value:", pval)

if pval < 0.05:
    print("✅ Significant correlation between Rat arrivals and Bat landings (p < 0.05)")
else:
    print("❌ No significant correlation (p >= 0.05)")
```

```
Contingency Table (Risk vs Reward):

reward    0    1
risk
0        72  386
1       351   98

Chi-square Test Results (Risk vs Reward):
Chi2 statistic: 352.83381148693195
Degrees of freedom: 1
Expected frequencies:
 [[213.59867696 244.40132304]
 [209.40132304 239.59867696]]
p-value: 1.0233966969013452e-78

✅ Significant relationship between Risk and Reward (p < 0.05)

Correlation Test (Rat arrivals vs Bat landings):
Correlation coefficient (r): -0.07353632546043398
p-value: 0.0006967431337288468
✅ Significant correlation between Rat arrivals and Bat landings (p < 0.05)
```

Conclusion

From the descriptive and inferential analysis:

Risk Behaviour: The majority of bats showed risk avoidance, but a portion engaged in risk taking.
Reward Behaviour: Reward outcomes varied, suggesting risk-taking does not always guarantee food access.
Seasonal Patterns: Risk-taking appeared to differ across seasons, possibly influenced by food availability.
Rat vs Bat Relationship: Scatterplots and correlation tests indicated whether increased rat arrivals reduce bat landings.

Interpretation:
If bats consistently avoid feeding when rats are present, it supports the hypothesis that bats perceive rats as predators (not just competitors).
If there is no significant difference, then rats may be seen mainly as competitors.

This evidence contributes to Investigation A and addresses Objective 1 (Assessment 2).