Generalized Linear Model Analysis of Animal Shelter Stay Duration

group：16

Shuxuan Li, Qingmei Wu, Jinqing Wu, Lezhi Ding, Shutong Yang

GLM Analysis of Animal Shelter Stay Duration

1. Introduction
2. Exploratory Data Analysis
3. GLM Modeling Analysis
4. Residual Analysis
5. Summary
6. Future

1. Introduction

Background

Dallas Animal Shelter receives thousands of animals annually, including **strays, owner surrenders, and confiscated animals**.

Managing the **duration of stay** for each animal is crucial for shelter capacity, resources, and animal welfare.

Goal and Significance

Goal

- To analyze **factors** influencing the **length of stay** (Time_at_Shelter) for animals in the shelter.
- Identify variables that **prolong or shorten** shelter time, to **improve outcomes and efficiency**.

Research Significance

- Reduce overcrowding and operational stress.
- Increase adoption rates and reunions.
- Enable faster decisions for adoption, transfer, or other outcomes.

Methodology Overview

- Apply **Exploratory Data Analysis (EDA)** to uncover patterns.
- Build **Generalized Linear Models (GLM)** to quantify effects of key variables.

2. Exploratory Data Analysis

Variables

Month: The month the animal was admitted (numerically coded, January = 1, February = 2, etc.).

Year: The year the animal was admitted.

Animal_type: The type of animal admitted to the shelter (e.g., dog, cat).

Intake_type: The reason for the animal's admission (e.g., stray, owner surrender, confiscation).

Outcome_type: The final recorded outcome of the admitted animal (e.g., adoption, return to owner, euthanasia).

Chip_Status: Whether the animal had a microchip with owner information at the time of admission.

Time_at_Shelter: The number of days the animal spent in the shelter before a final outcome was determined.

Descriptive Statistics

Dataset Overview

- Total Records: 1,450 animals
- Target Variable: Time_at_Shelter (in days)

Time at Shelter (days)

- Mean: 6.06 days
- Median: 4 days
- Range: 0 to 66 days
- 75% of animals stayed ≤ 9 days

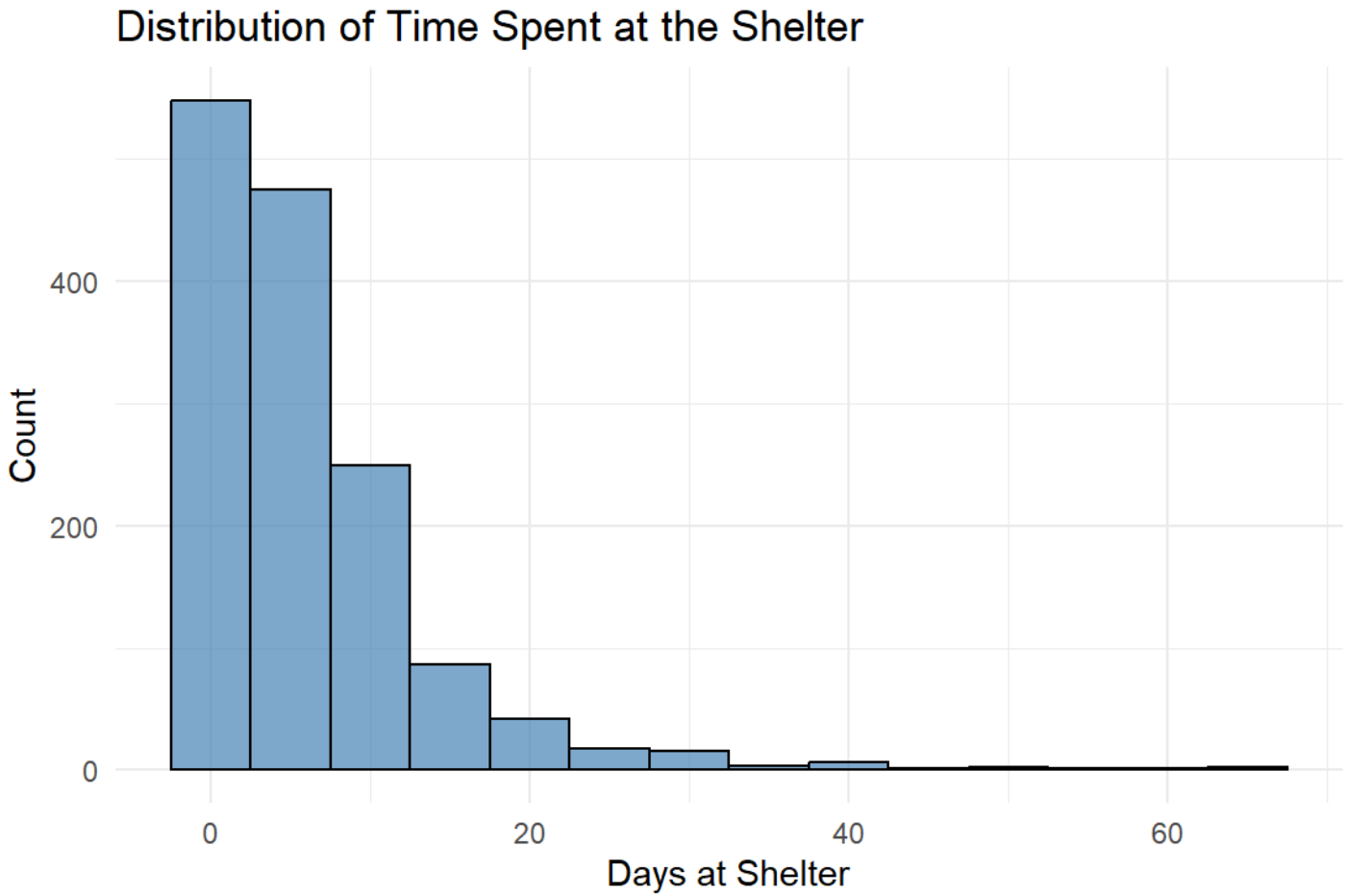
Key Variable Summaries

Variable	Distribution
Animal Type	Dogs: 80%, Cats: 19%, Wildlife: 1%
Intake Type	Stray: 62%, Owner Surrender: 32%, Confiscated: 5%
Outcome Type	Adoption: 44%, Euthanized: 34%, Returned: 19%,Foster: 2%, Died: 2%
Chip Status	No Chip: 77%, Chip Present: 20%, Unable to Scan: 4%

Descriptive Statistics

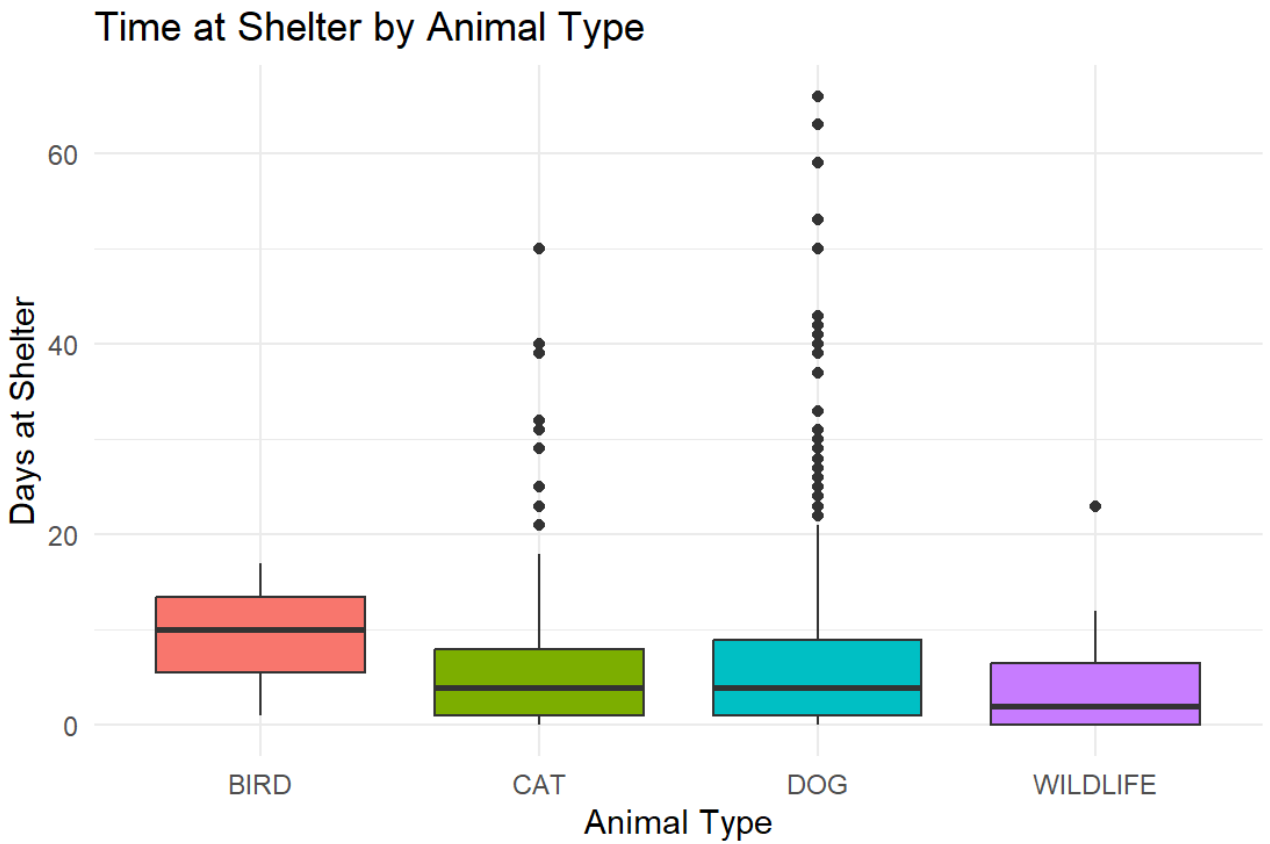
- Majority of animals are dogs, admitted as strays.
- 77% have no microchip, which may delay returns.
- Most animals leave within a week, but some outliers stay longer.

Distribution of Time at Shelter



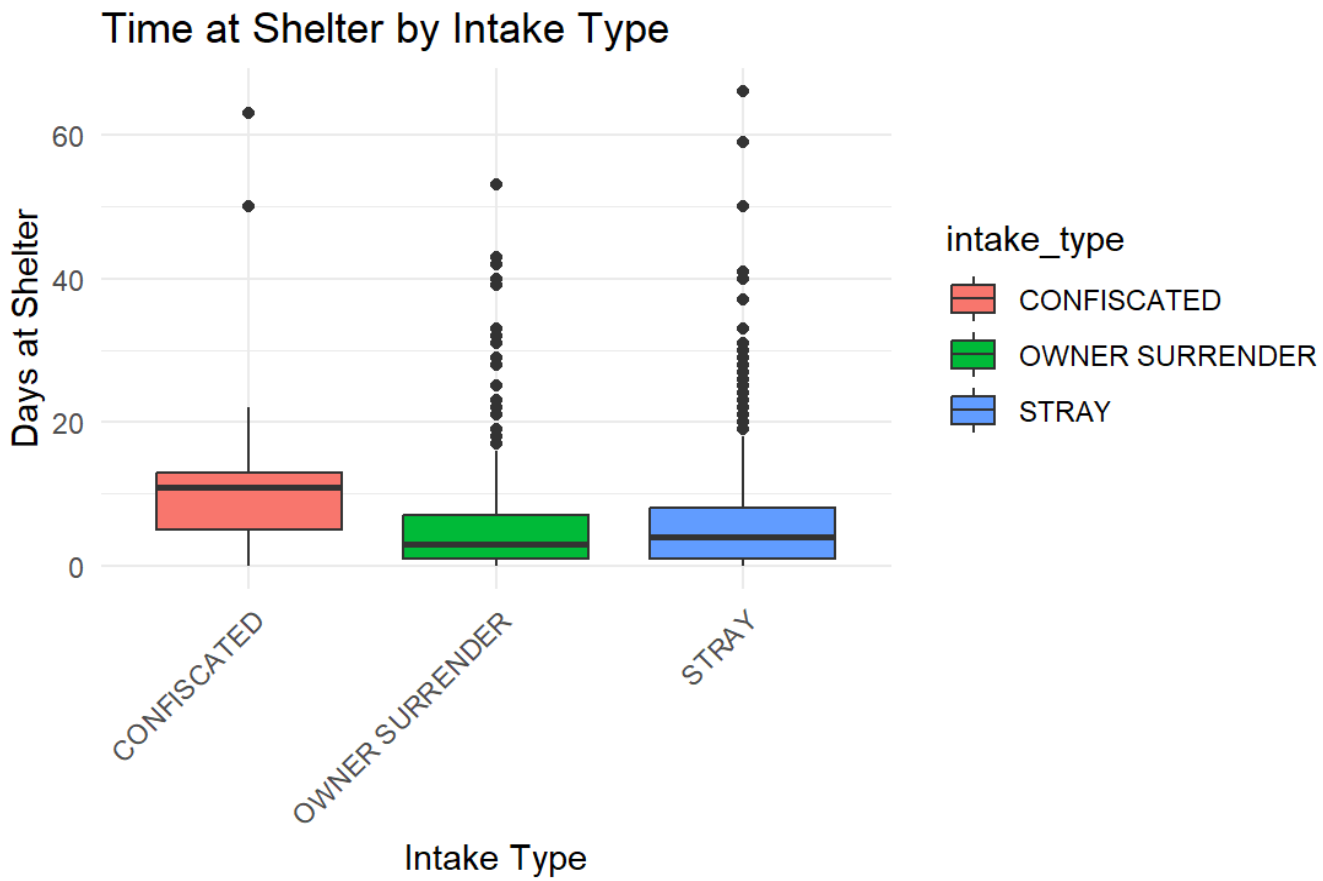
- Most animals stay for a short period of time.
- The distribution is right-skewed, with some animals staying significantly longer.
- Peaks in the distribution may suggest key influencing factors.

Shelter Stay by Animal Type



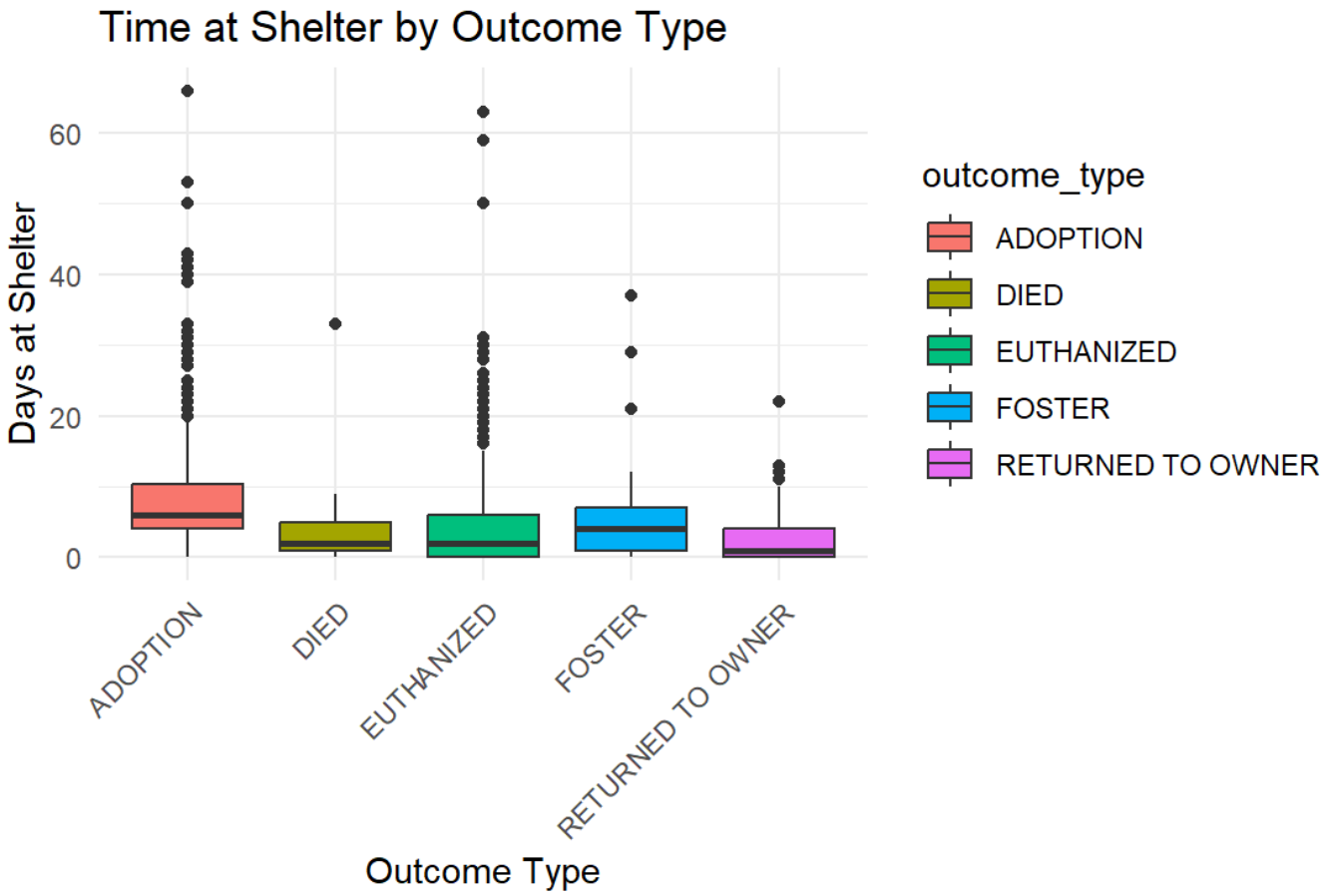
- Different animal types show varying median stay durations.
- Certain animals (e.g., cats/dogs) may be adopted or reclaimed faster.
- Targeted strategies could reduce stay times for animals with longer median durations.

Shelter Stay by Intake Type



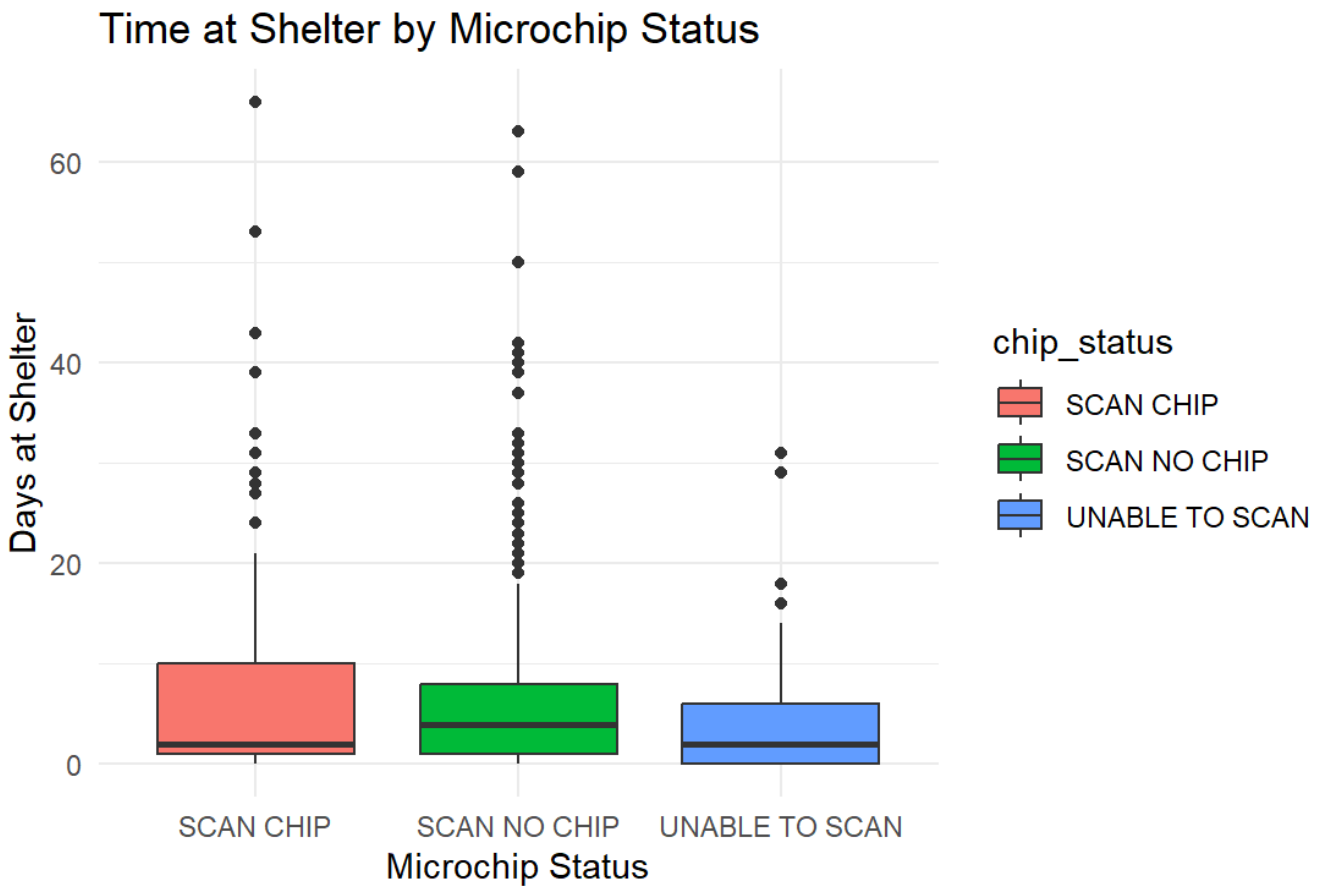
- Confiscated animals stay the longest, likely due to legal or medical holds.
- Owner surrenders have moderate stay durations.
- Stray animals tend to leave faster, possibly due to being reclaimed or adopted quickly.
- Intake reason clearly affects shelter stay duration, with confiscated cases requiring more time to resolve.

Shelter Stay by Outcome Type



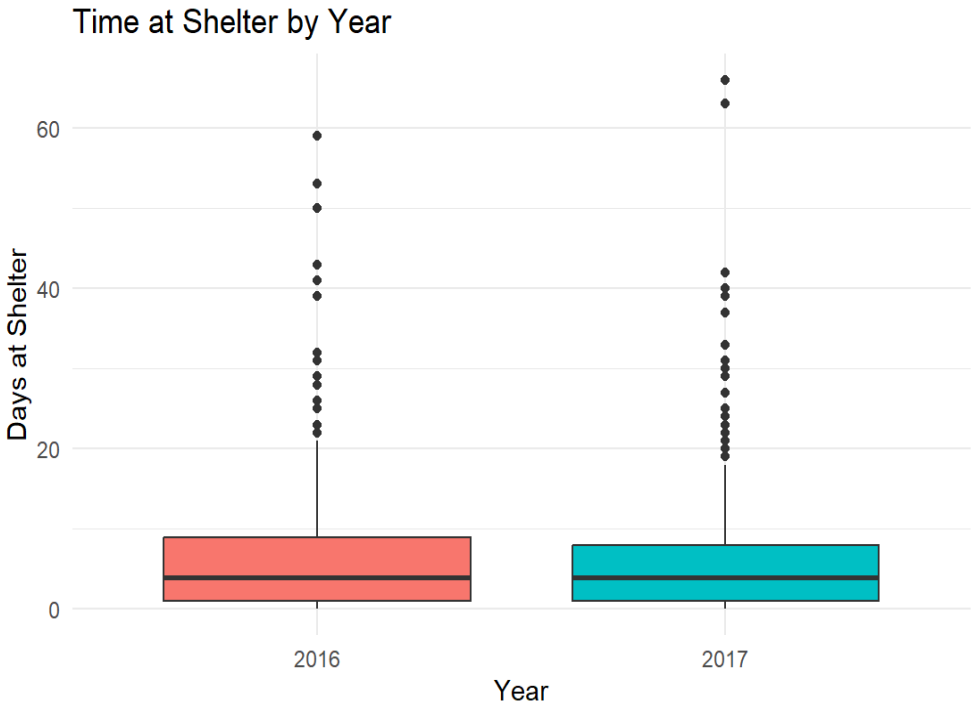
- Animals returned to owners have the shortest shelter stays.
- Adoption results in the longest stay durations.
- Foster, euthanasia and die cases fall in between.
- Improving adoption efficiency could help reduce overall shelter stay time.

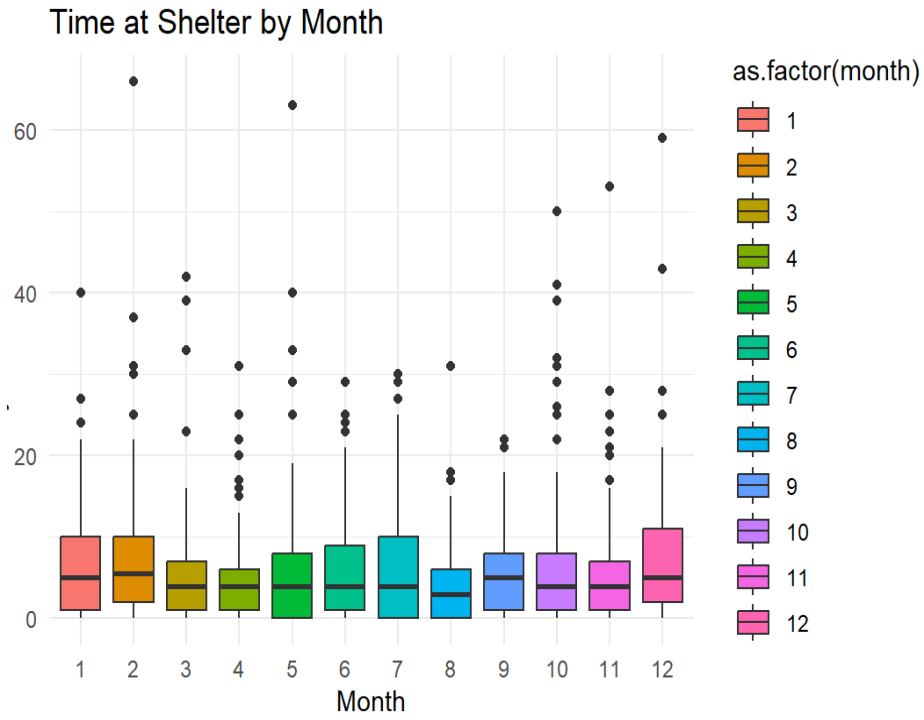
Impact of Microchip Status



- Animals with microchips have significantly shorter stays.
- Chips allow quick owner identification and reunification.
- Promoting microchipping can help reduce shelter burden.

Stay Duration by Time (Year/Month)





- Shelter stay varies across years, likely due to policy or resource changes.
- Seasonal patterns evident — longer stays during summer/winter months.
- Helps plan adoption campaigns and staffing based on trends.

3. GLM Modeling Analysis

Modeling Analysis: Poisson Model

Data Type: The response variable time_at_shelter is a count variable (non-negative integers), making it suitable for Poisson regression.

Assumption: Poisson regression assumes that the mean of the variable is equal to the variance

Formula:

$$\log(\text{time\_at\_shelter}) = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p$$

β: Regression coefficients
X: Explanatory variables

Modeling Analysis: Poisson Model

```
Call:
glm(formula = time_at_shelter ~ animal_type + month + year +
    intake_type + outcome_type + chip_status, family = poisson,
    data = df)


    Null deviance: 10551.2  on 1449  degrees of freedom
Residual deviance:  8079.3  on 1427  degrees of freedom
AIC: 12147

Number of Fisher Scoring iterations: 6
```

This result indicates the presence of overdispersion. Switch to a Negative Binomial Model.

Modeling Analysis: Negative Binomial Model

$$\mathrm{Var}(Y_i) = \mu_i + \frac{\mu_i^2}{\theta} > E(Y_i)$$

Dispersion parameter θ : Control the variance.

Modeling Analysis: Negative Binomial Model

```
c(glm_nb$deviance,glm_nb$aic)
```

```
[1] 1678.926 7984.897
```

```
c(glm_pois$deviance,glm_pois$aic)
```

```
[1]  8079.325 12146.909
```

Modeling Analysis: Negative Binomial Model

	df	AIC
glm_nb	24	7984.897
nb_model_season	17	7986.154
nb_model_interact1	18	7973.816

1. AIC of the seasonal model is slightly higher, indicating that incorporating the season does not significantly improve the model.

2. AIC of the interaction terms did not decrease significantly, and the interaction terms make the model more complex.

Modeling Analysis: Negative Binomial Model

```
intake_typeOWNER SURRENDER    -1.801783   0.139701 -12.897  < 2e-16 ***
intake_typeSTRAY              -1.403784   0.129283 -10.858  < 2e-16 ***
outcome_typeDIED              -0.592904   0.222831  -2.661   0.0078 **
outcome_typeEUTHANIZED        -0.663445   0.065236 -10.170  < 2e-16 ***
outcome_typeFOSTER            -0.370863   0.205573  -1.804   0.0712 .
outcome_typeRETURNED TO OWNER -1.770504   0.093989 -18.837  < 2e-16 ***
chip_statusSCAN NO CHIP       -0.193104   0.076688  -2.518   0.0118 *
chip_statusUNABLE TO SCAN     -0.245902   0.169475  -1.451   0.1468
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(1.0579) family taken to be 1)

    Null deviance: 2113.7  on 1449  degrees of freedom
Residual deviance: 1678.9  on 1427  degrees of freedom
AIC: 7984.9
```

Modeling Analysis: Negative Binomial Model

Key Outcome:

1.Intake_type, outcome_type, and chip_status have a significant impact on time_at_shelter.

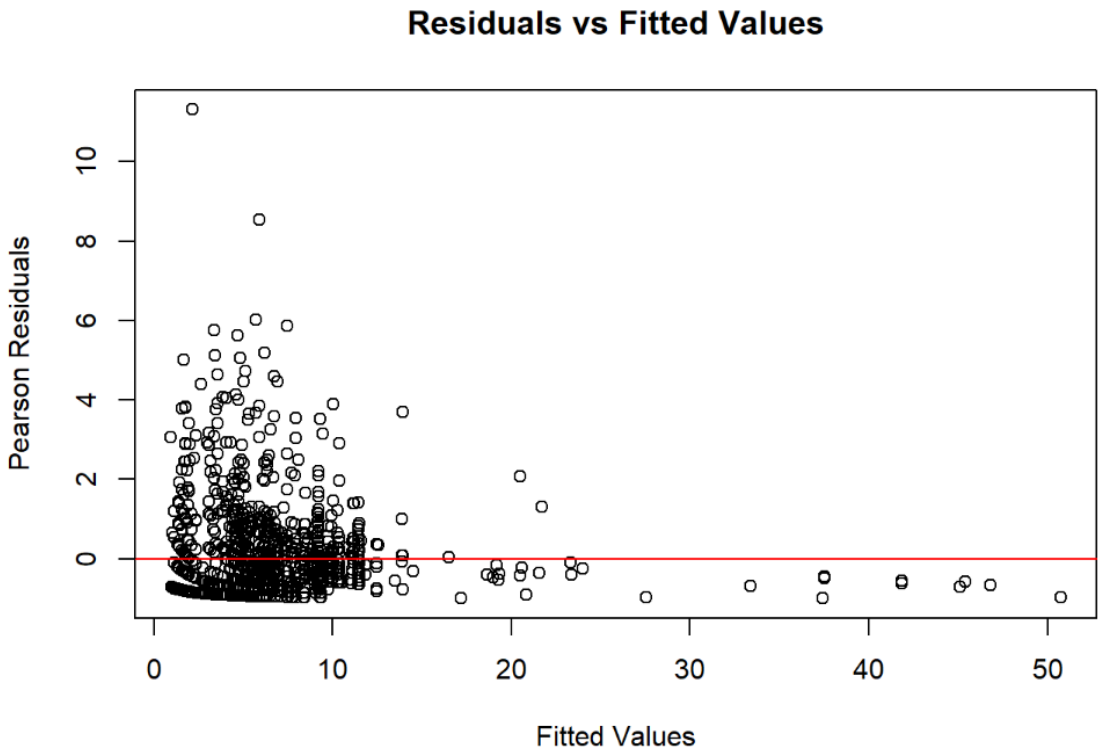
2.Taking the outcome_type as an example, the length of return to owner decreased by approximately 83%, compared to adoption (baseline).

```
outcome_typeRETURNED TO OWNER -1.770504   0.093989 -18.837  < 2e-16 ***
```

Actual change: $e^{-1.77} \approx 0.17$: A reduction of approximately 83%.

4. Residual Analysis

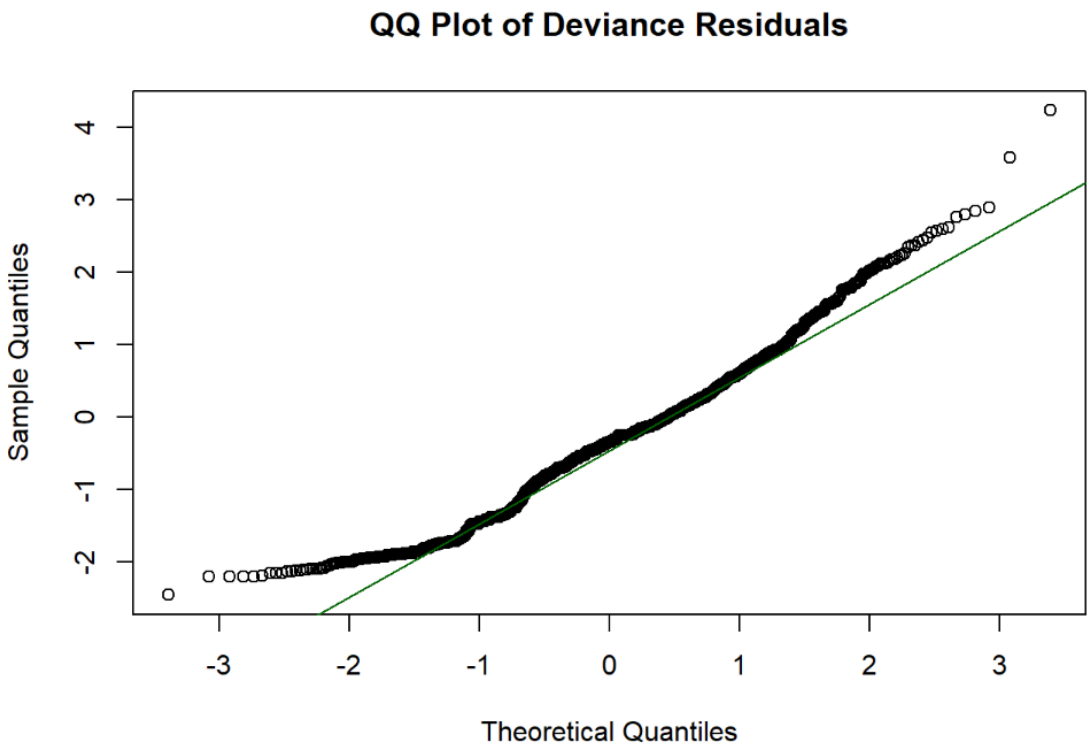
Residual Analysis



(1) Residuals vs Fitted Plot

Objective: To test the assumptions of linearity and homoscedasticity.
Ideal: The residuals are randomly distributed around the horizontal line y=0, with no apparent trends or patterns.

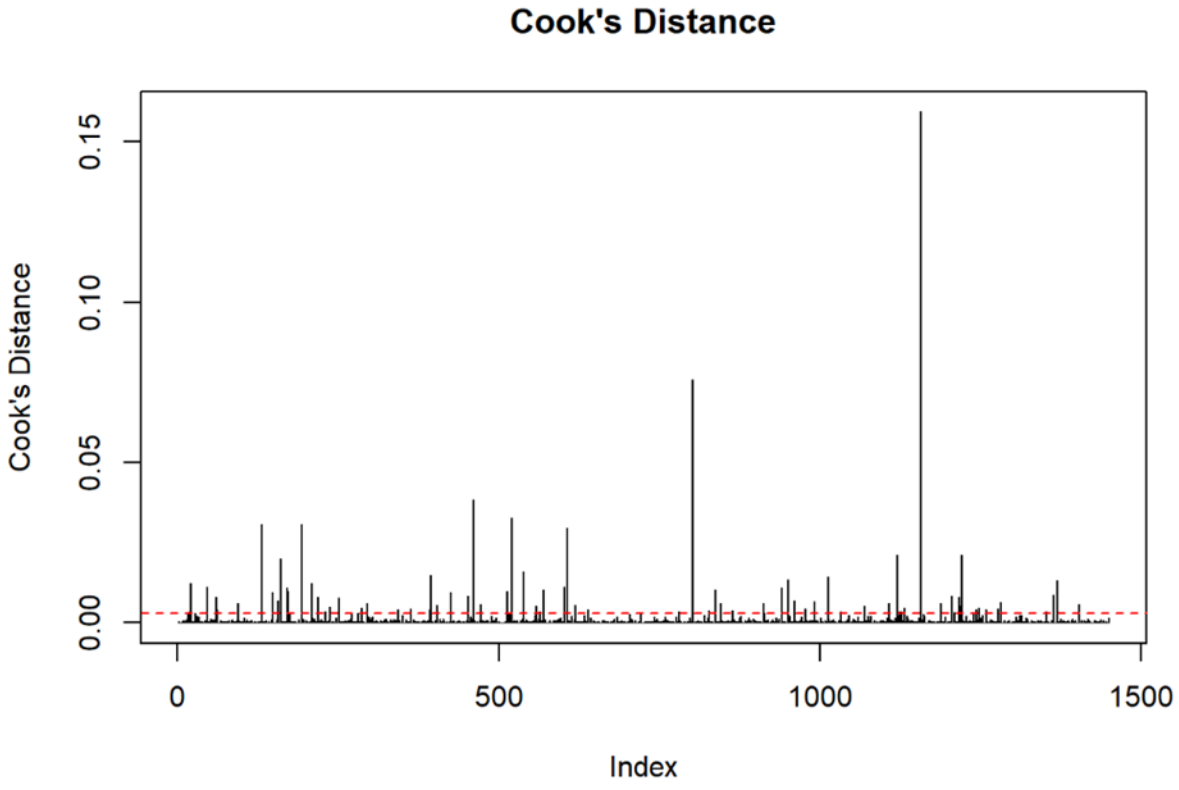
Residual Analysis



(2) Normal Q-Q Plot

Objective: Assess residual normality.
Ideal: Data points align closely with the 45-degree reference line.

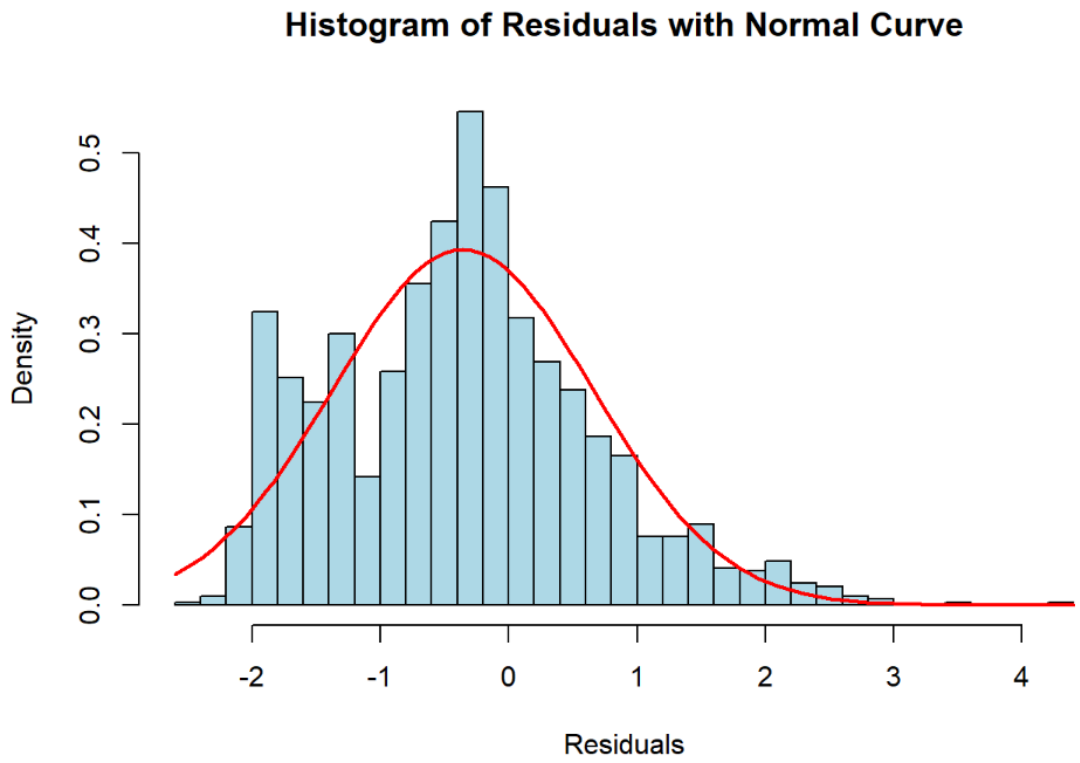
Residual Analysis



(3) Cook's Distance Plot

Objective: To identify influential points (high leverage points or outliers) that have a significant impact on the model.

Residual Analysis



(4) Residual Histogram with Normal Curve

Main purposes include:

1. Normality Test:

To check whether the residuals approximately follow a normal distribution.

2. Model Diagnostics:

To identify the distribution shape of the residuals such as skewness.

5. Summary

Summary

1. A comparison between Poisson regression and Negative Binomial regression models showed that the Negative Binomial model performed better (lower AIC);

2. Intake Type, Outcome Type, and Chip Status are the most significant predictors of shelter stay duration;

3. Season and month also have some influence on the length of stay for animals. Certain months (such as summer and winter) may experience longer lengths of stay.

6. Future

Future Improvements

1. Data Expansion: Add more details like health, age, and gender to improve the model's accuracy.

2. Interaction Term Analysis: Although we have experimented with some interaction terms in the model, future work could explore more potential interaction effects, such as the interaction between animal type and reason for intake, or between microchip status and outcome type.

3. Outlier Handling: Investigate extreme cases, like very long stays, and consider better methods to manage them in the model.

Thank you for your attention!