Analysis of the number of days an animal spends in the shelter

Group17

Reporter: Jiaxiang LI ;Yutong LU ; Yunxi ZHANG

University of Glasgow

CONTENTS

Aims of the analysis

- **Aim:**
- Research into the number of days an animal spends in the shelter is expected to help shelters make better decisions.

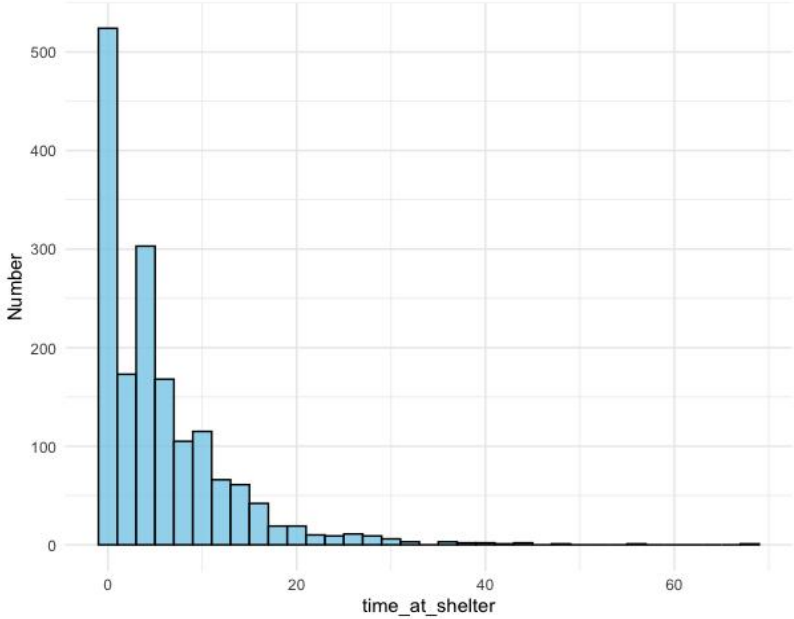
- **Problem Statement:**
- Which factors influence the number of days an animal spends in the shelter before its final outcome is decided?

Data Overview

Key variables:

- Animal_type – The type of animal admitted to the shelter
- Month – Month the animal was admitted, recorded numerically with January=1
- Year. – Year the animal was admitted to the shelter.
- Intake_type – Reason for the animal being admitted to the shelter
- Outcome_type – Final outcome for the admitted animal
- Chip_Status – Did the animal have a microchip with owner information?
- Time_at_Shelter – Days spent at the shelter between being admitted and the final outcome.

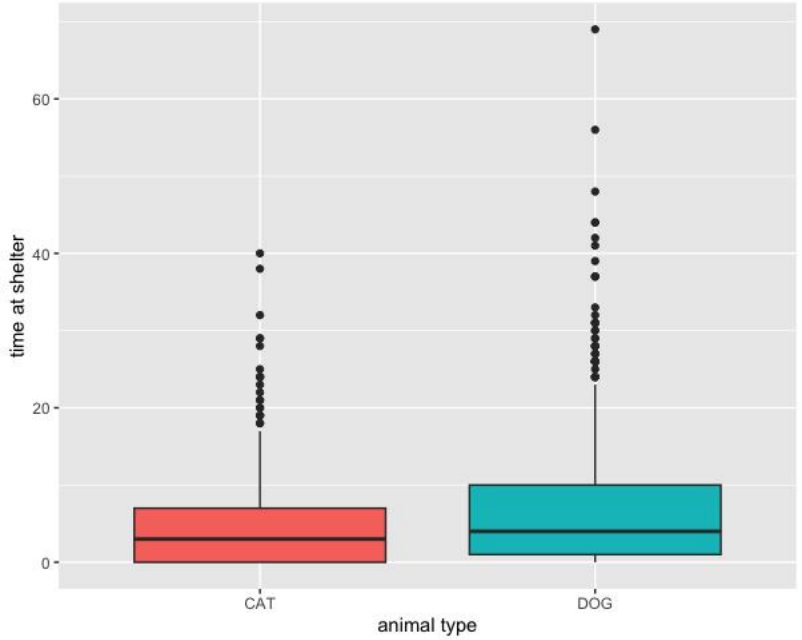
1. **Histogram** of the number of days an animal spends in the shelter



The chart shows a right-skewed distribution, with most animals staying 0-10 days and few staying beyond 30 days.

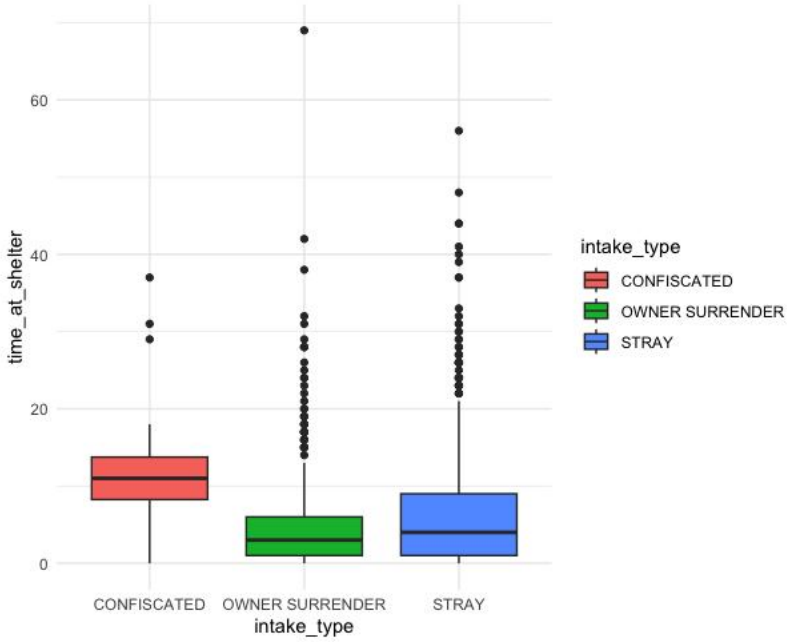
Data Overview

2. **Boxplot** of time_at_shelter and animal_type



Cats have a lower median stay than dogs, but dogs show a wider distribution with longer stays.

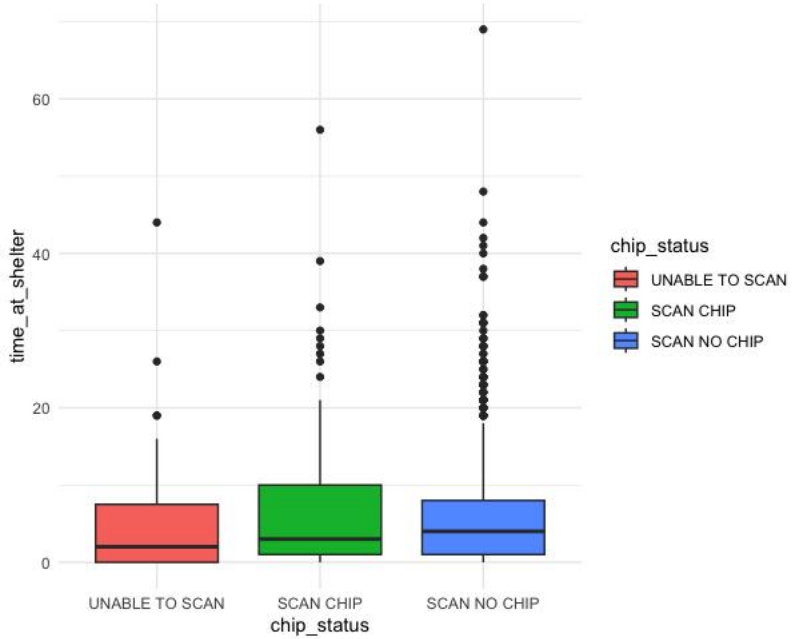
3. **Boxplot** of time_at_shelter and intake_type



The median of confiscated animals was significantly higher than that of the other two groups, indicating that confiscated animals stayed concentrated and longer.

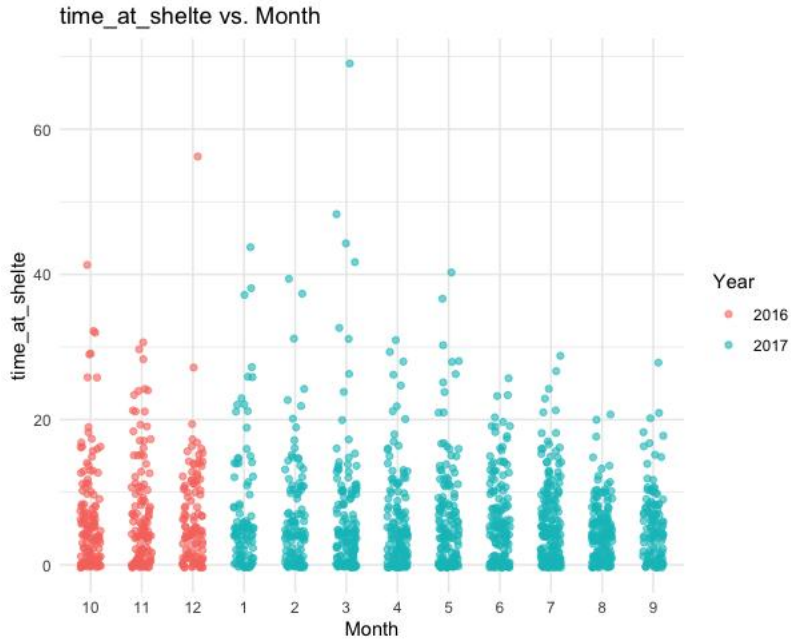
Data Overview

4. **Boxplot** of time_at_shelter and chip_status



The medians were similar across chip statuses, suggesting comparable shelter stays. However, the IQR was wider for scanned animals, indicating greater data dispersion.

5. **Scatter plot** of time_at_shelter and intake_type



The dataset spans from October 2016 to September 2017 with complete monthly data. Animal shelter stays were mostly short (under 20 days) and showed no significant seasonal variation.

Poisson Regression

	Pr
(Intercept)	< 2e-16 ***
animal_typeDOG	4.70e-09 ***
month_ordered	< 2e-16 ***
intake_typeOWNER SURRENDER	< 2e-16 ***
intake_typeSTRAY	< 2e-16 ***
chip_statusSCAN CHIP	6.58e-07 ***
chip_statusSCAN NO CHIP	7.55e-12 ***

AIC: 15692

Serious ***Overdispersion***!

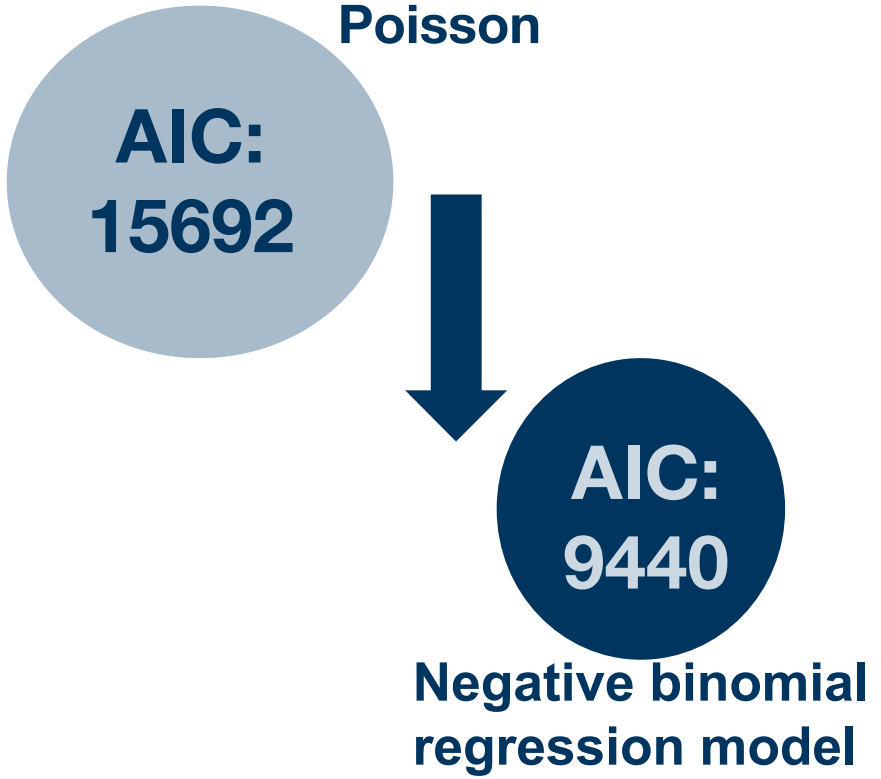
Dropped

Dispersion 8.01>>1.5

Negative binomial regression model

	Pr
(Intercept)	< 2e-16 ***
intake_typeOWNER SURRENDER	2.10e-09 ***
intake_typeSTRAY	1.88e-05 ***
month_ordered	7.66e-05 ***
chip_statusSCAN CHIP	0.03571 *
chip_statusSCAN NO CHIP	0.00491 **
animal_typeDOG	0.09949 .

Stepwise used to select optimal variables

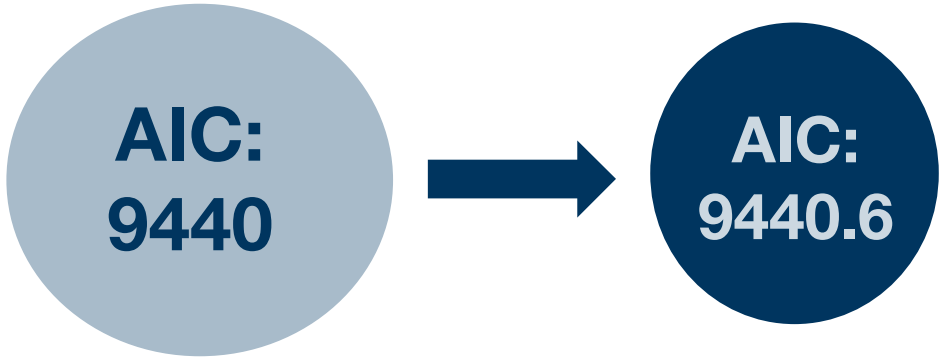


Negative binomial regression model

	Pr
(Intercept)	< 2e-16 ***
intake_typeOWNER SURRENDER	3.68e-10 ***
intake_typeSTRAY	1.27e-05 ***
month_ordered	5.00e-05 ***
chip_statusSCAN CHIP	0.01969 *
chip_statusSCAN NO CHIP	0.00323 **

After excluding animal_type, it performs similarly but is simplified

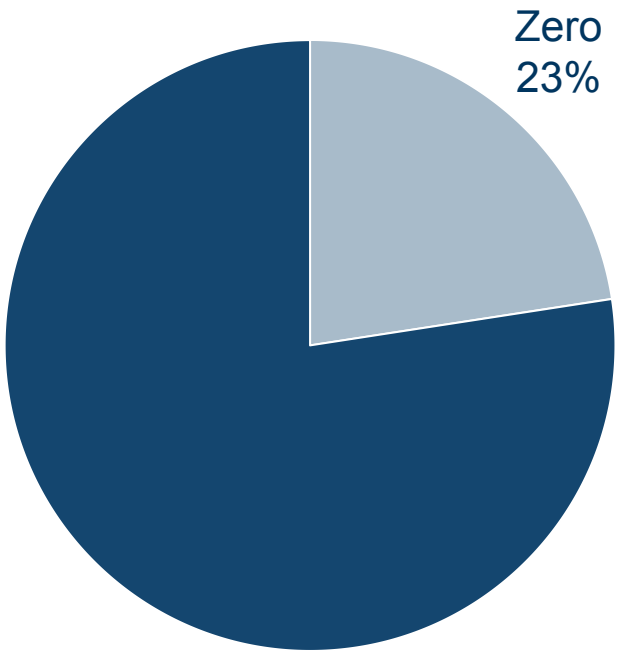
include animal_type



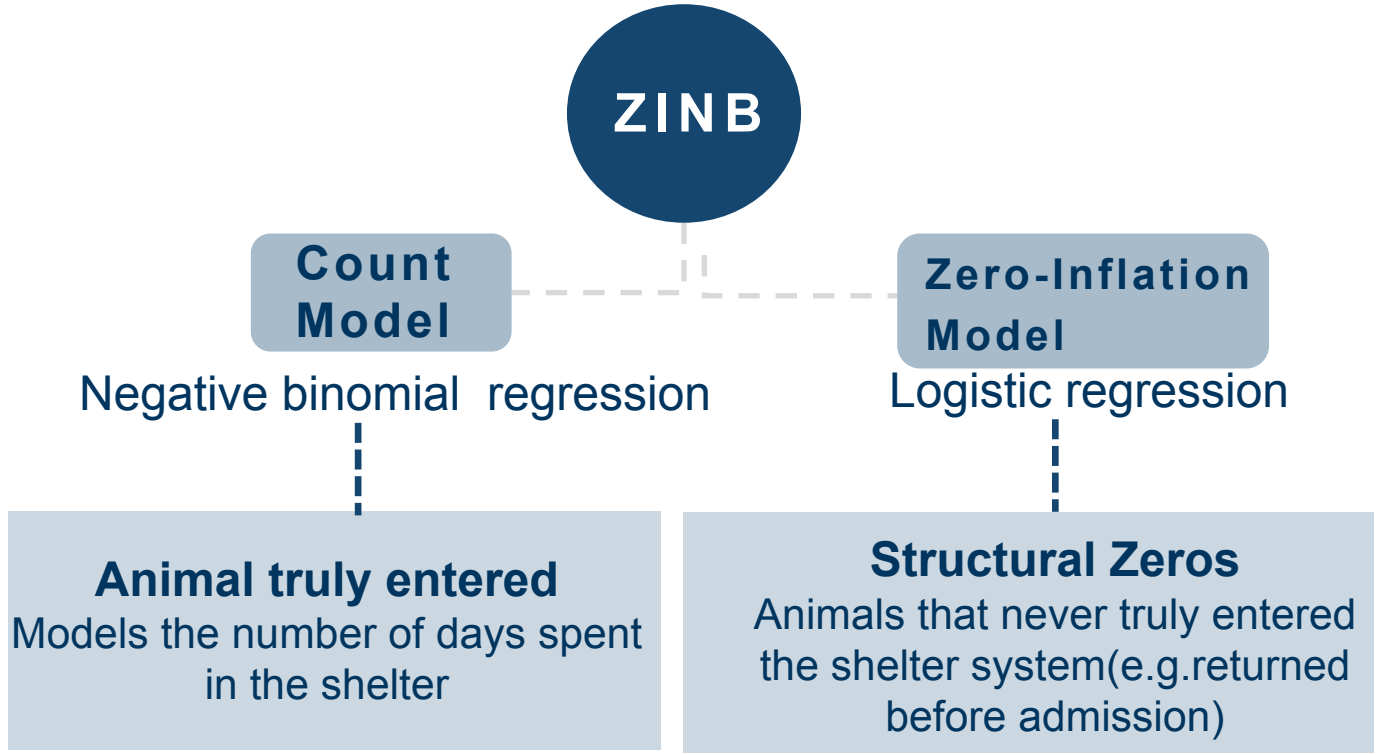
Exclude animal_type

Zero-Inflated Negtive Binomial Model

The amount of Zero in *time_at_shelter*



Excess zeros in *time_at_shelter*



Zero-Inflated Negtive Binomial Model

Count model

Estimates how many days an animal stays in the shelter (if it truly entered)

	Estimate	Pr
(Intercept)	2.513081	< 2e-16 ***
intake_typeOWNER SURRENDER	-0.657431	6.47e-09 ***
intake_typeSTRAY	-0.345890	0.001047 **
month_ordered	-0.026263	0.000668 ***
chip_statusSCAN CHIP	-0.012541	0.938263
chip_statusSCAN NO CHIP	0.033997	0.823711
animal_typeDOG	0.030271	0.664766
Log(theta)	0.359465	4.38e-08 ***

AIC = 9335.147 lowest among three model

OWNER SURRENDER and ***STRAY*** have significant negative coefficients → These animals tend to stay fewer days. Stray animals stay exp(-0.346) ≈ 0.71 times as long.

month_ordered is also significantly negative → animals arriving later in the year tend to stay shorter. For each later month, time in shelter decreases by 2.7%.

chip_status and ***animal_typeDOG*** are not significant → have little impact on length of stay.

Zero-Inflated Negtive Binomial Model

Zero-inflation model

Models the probability that a zero is a structural zero – animal didn't really enter the shelter

	Estimate	Pr
(Intercept)	-12.35866	0.923125
intake_typeOWNER SURRENDER	12.14699	0.924438
intake_typeSTRAY	12.36739	0.923071
month_ordered	0.04147	0.106656
chip_statusSCAN CHIP	-1.36654	0.000202***
chip_statusSCAN NO CHIP	-1.59158	3.01e-07 ***
animal_typeDOG	-0.49316	0.016035 *

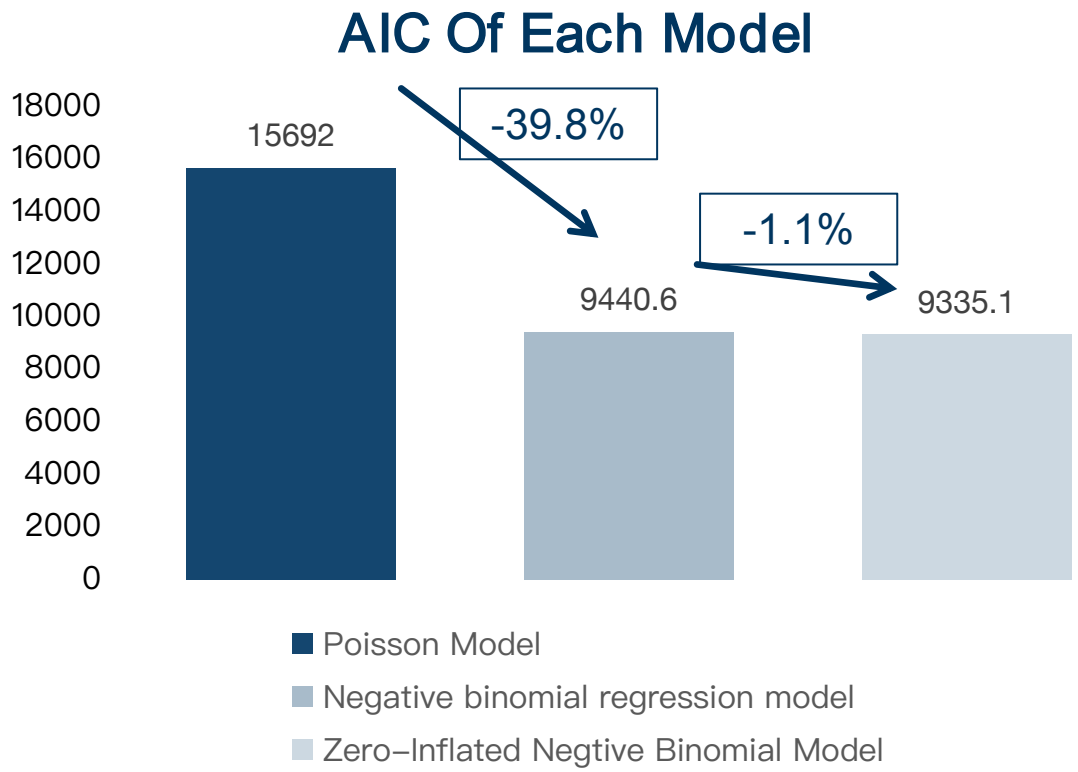
AIC = 9335.147 lowest among three model

SCAN CHIP and ***SCAN NO CHIP*** have significant negative coefficients → Animals with chip-related info are less likely to be structural zeros. Having a scannable chip ***reduces odds*** of being a structural zero by exp(-1.37) ≈ 0.25.

animal_typeDOG is significant and negative → Dogs are less likely to be structural zeros. Dogs are exp(-0.49) ≈ 0.61 times as likely as cats to be structural zeros.

intake_type and ***month_ordered*** are not significant here. → Have little impact on structural zero.

Model Comparison

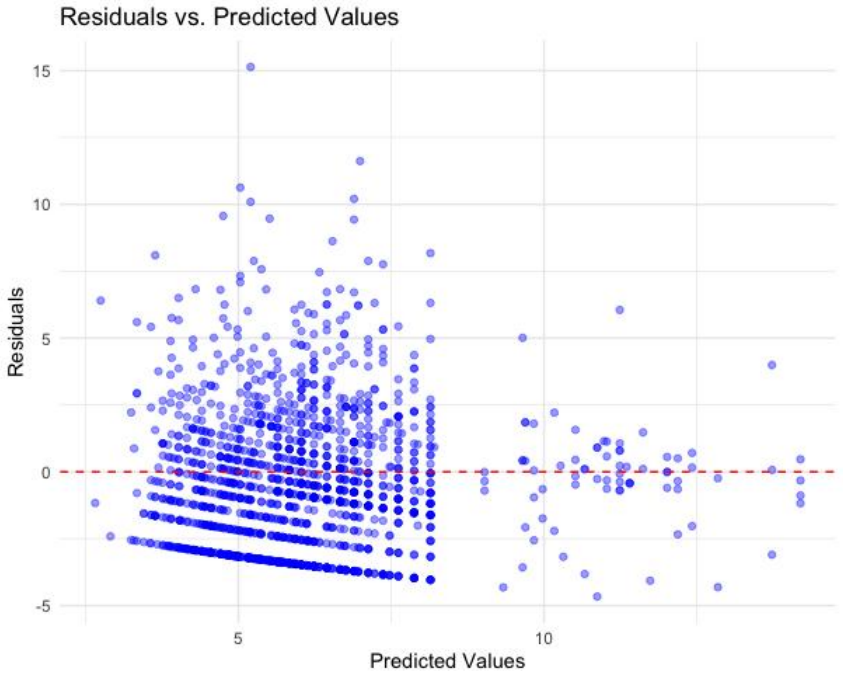


The Zero-Inflated Negative Binomial (ZINB) model has the lowest AIC of 9335, indicating the best fit.

It captures both overdispersion and excess zeros, which the Poisson and NB models cannot fully handle.

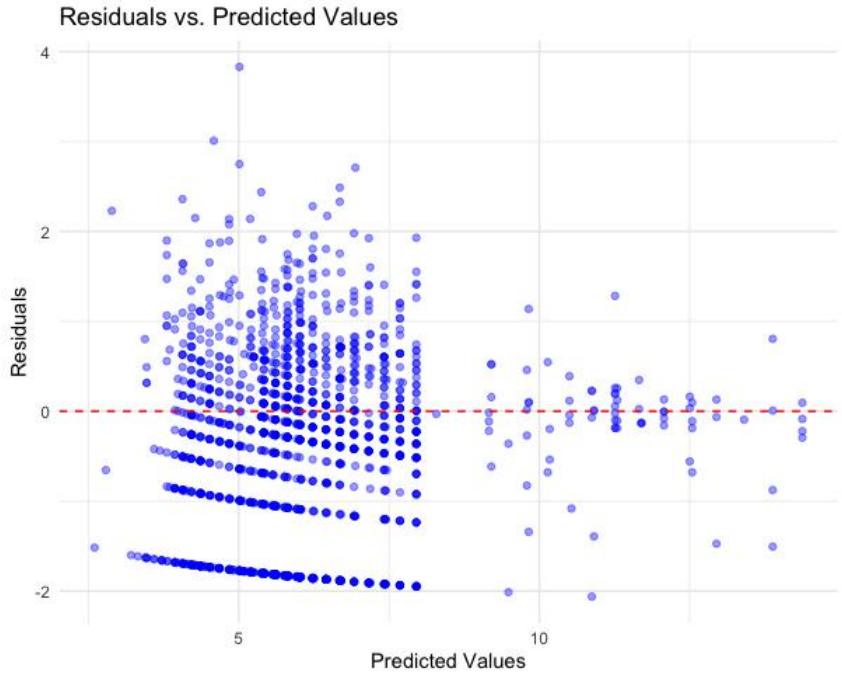
Residual Plot

Model1 :
Possion



Residual variability is higher at lower predicted values, indicating excessive dispersion, consistent with the previously calculated dispersion parameter.

Model2 :
nb_model_2

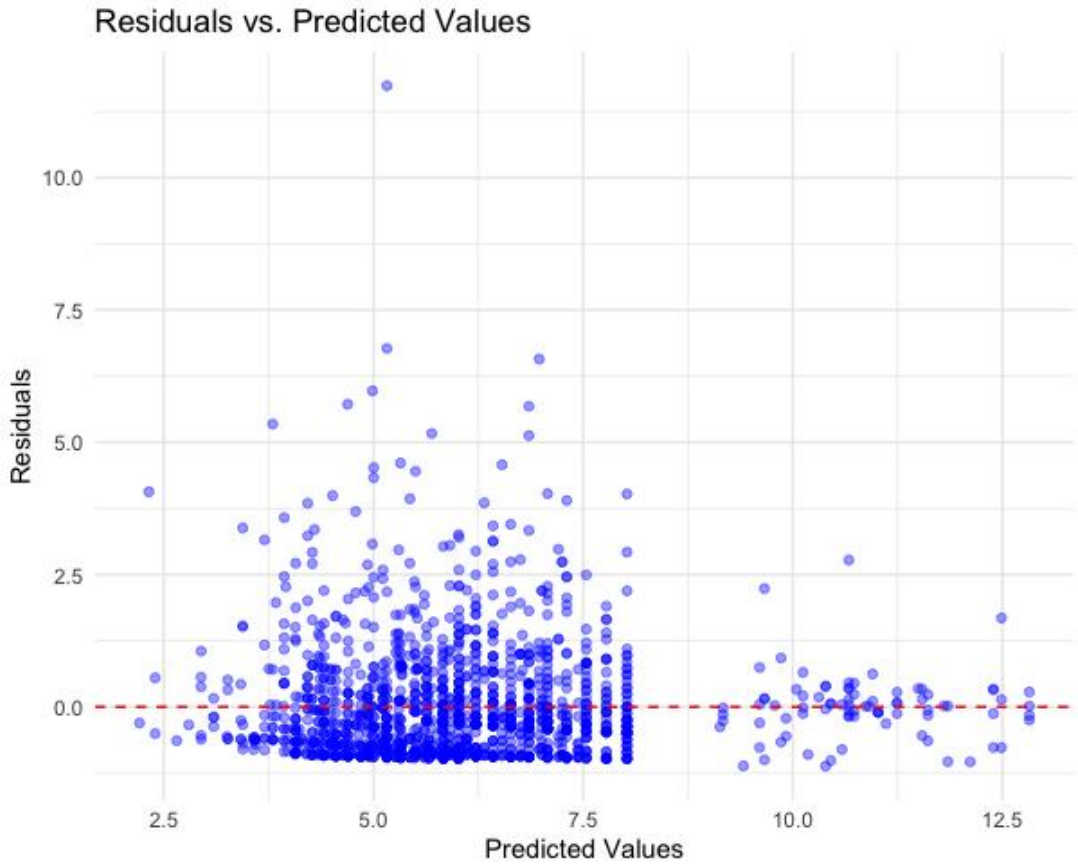


The variance of the residuals is generally in the range of -2 to +2. The negative binomial regression model basically solves the overdispersion problem.

Residual Plot

Model3 :
Zero-flated_model

The residual plot highlights useful insights, confirming the presence of dispersion and validating the need for a more flexible model.



Key Findings Summary

Distribution

This report examines animal shelter stays, showing a right-skewed distribution with most animals staying under 10 days.

Predictors

Using Poisson and Negative Binomial models with predictors like intake_type, animal_type, chip_status, and month, we found that owner-surrendered or stray animals stay shorter, dogs stay slightly longer, and chip status and intake month have weaker effects.

Modeling

Given many zero-day stays, a Zero-Inflated Negative Binomial model provided the best fit, distinguishing structural zeros from regular counts for deeper insight.

Limitations & Future Work

Problem statment

- The dataset contains a large number of zero values.
- Multiple explanatory variables are present.
- GLM does not fit well.

Proposed solution

- Explore ANN as an alternative predictive model.
- Investigate methods for transforming non-numerical variables into numerical ones.
- Consider whether the frequency of observer occurrences can serve as a conversion criterion.


University of Glasgow
THANKS!