Group_16_Project

Group_16

Introduction

Animal protection has now become a growing concern in society and animal being admitted to the shelter is one of the conservation measures.The purpose of this report is to use the Generalised Linear Model (GLM) to analyse whether the days of animals spent at the shelter between being admitted and the final outcome (Time_at_Shelter) is related to the type of animal (Animal_type),the month the animal was admitted (Month), the year the animal was admitted (Year), the reason for the animal being admitted (Intake_type), and status of a microchip with owner information (Chip_Status),and all the above will be anlyzed below.

Exploratory Data Analysis

In order to carry out an initial analysis of the relationship between the outcome variables and all the explanatory variables, we use five boxplots to provide an overview on a case-by-case basis.

For Figure 1,it shows that days spent at the shelter between being admitted and the final outcome for birds seem to be longer which means being with wider distribution.And there are more outliers for cats and dogs.

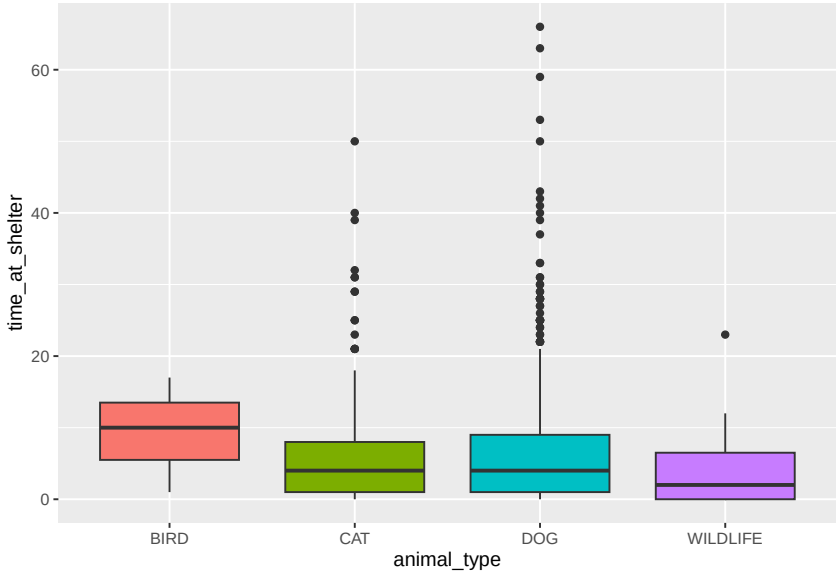

Figure 1: Figure 1:time_at_shelter by animal_type

Basically there is no difference of the days spent at the shelter between being admitted and the final outcome for different animals in 2016 and 2017 ,which can be showed in Figure 2.

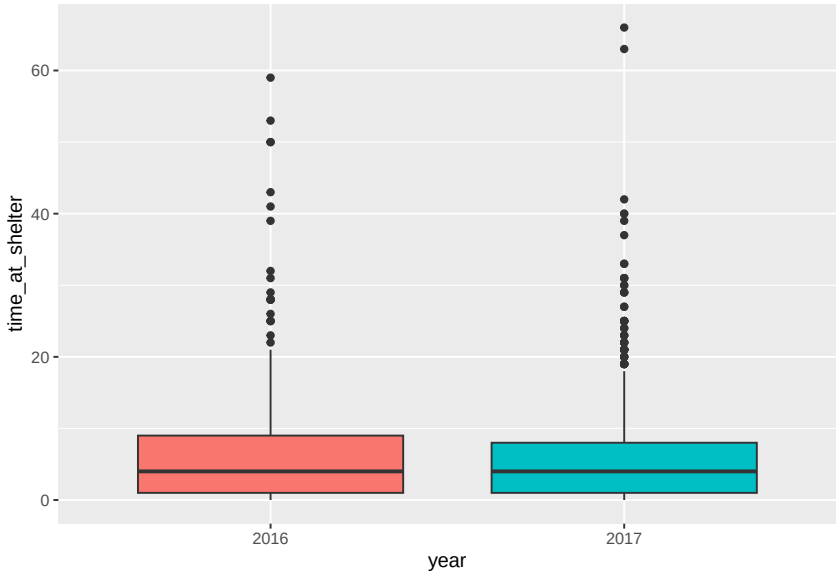

Figure 2: Figure 2:time_at_shelter by year

From Figure 3,we can see that for different months,the days spent at the shelter between being admitted and the final outcome for animals have their own differences,but the median for 12 months all seem to be similar.

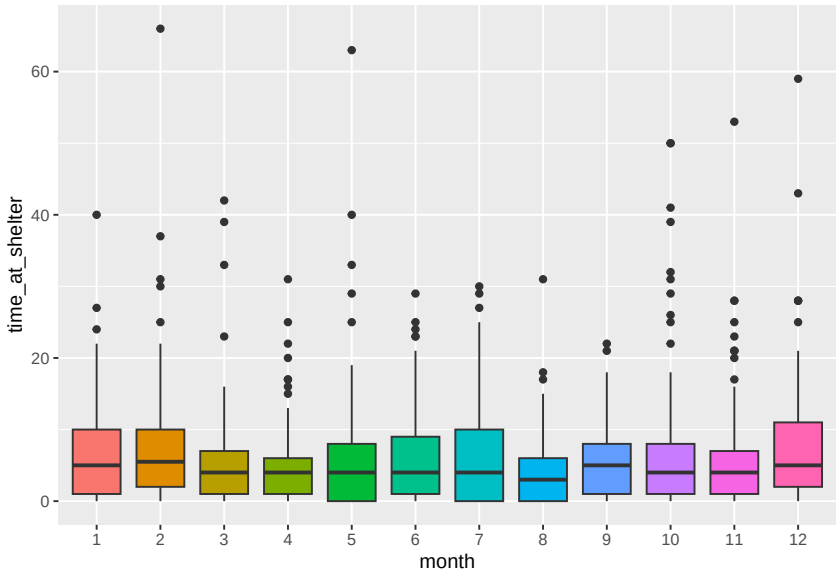

Figure 3: Figure 3:time_at_shelter by month

As is showed in Figure 4, the confiscated animals stay longest time in shelter before the outcome,and there are not too much difference for owner surrendered and stray animals.

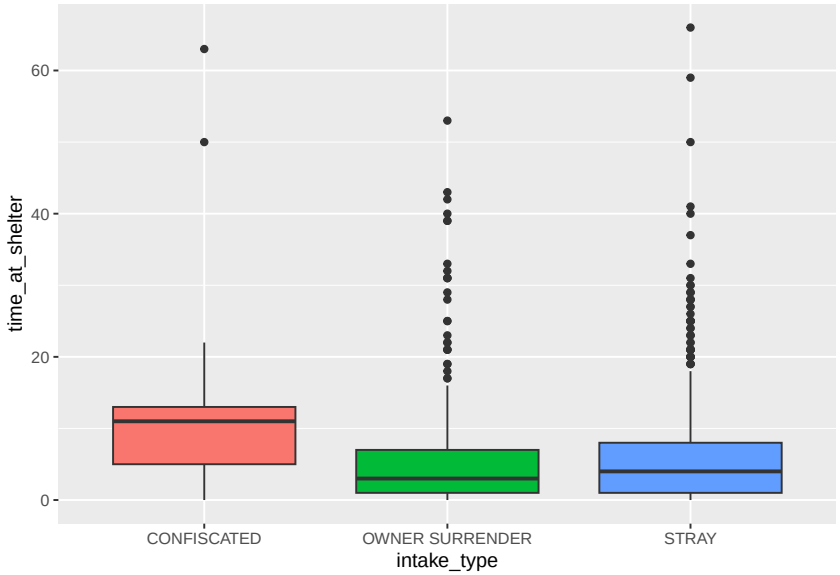

Figure 4: Figure 4:time_at_shelter by intake_type

In Figure 5,we can see that for the three types of chip status,the days in shelter for animals with scanned chips have smallest median but distribute more widely.

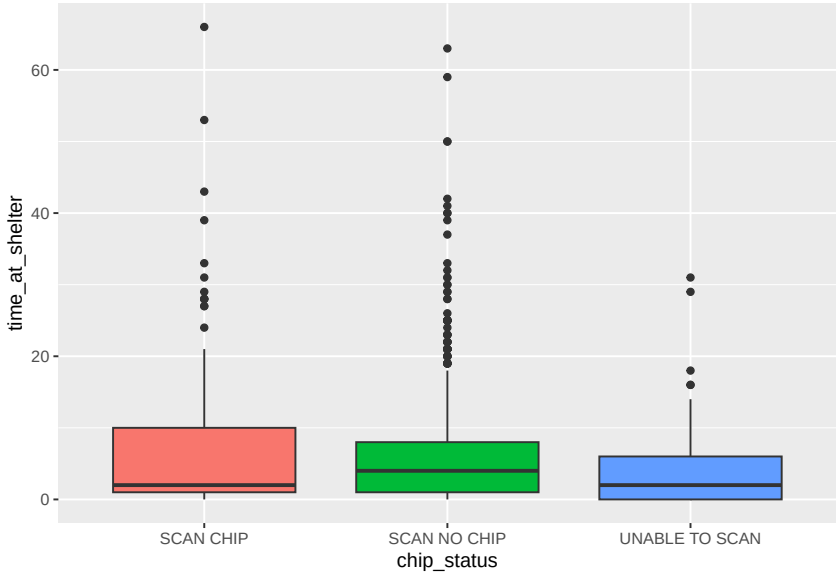

Figure 5: Figure 5:time_at_shelter by chip_status

Formal Data Analysis

After the initial graphical analysis of the relationship between the variables, we will use the GLM model to further analyse the relationship between the outcome variables (Time_at_Shelter)and the explanatory variables(Animal_type,Month,Year,Intake_type,Chip_Status), and the model will be based on poisson GLM model, since the response variable is the count of the number of days animal spent at the shelter, with link

function log.So,the model(1) will be like this:

$$\log(\mathrm{E}(y_i)) = \beta_0 + \beta_{1i}\mathrm{animal\_type_i} + \beta_{2j}\mathrm{month_j} + \beta_{3t}\mathrm{year_t} + \beta_{4p}\mathrm{intake\_type_p} + \beta_{5q}\mathrm{chip\_status_q}$$

where the value of $\mathrm{animal\_type_i}$ include:BIRD,CAT,DOG,WILDLIFE;the value of $\mathrm{month_j}$ include:1 to 12;the value of $\mathrm{year_t}$ include:2016,2017;the value of $\mathrm{intake_type_p}$include:CONFISCATED,OWNER SURRENDER,STRAY;the value of $\mathrm{chip\_status_q}$ include:SCAN CHIP,SCAN NO CHIP,UNABLE TO SCAN;and all the variables above are group variables.

The outcome can be showed as follow,and we notice that the coefficient of year factor cannot be defined.There are some reasons for that,firstly the the explanatory variables month and year exist multicollinearity,and there are also some missing month data in 2016,finally we can also find that changes in year does not affect the outcome variable from Figure 2.

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

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-4.9857  -2.6038  -0.7551   0.8060  12.7653

Coefficients: (1 not defined because of singularities)
                           Estimate Std. Error z value Pr(>|z|)
(Intercept)                 2.75306    0.19690  13.982  < 2e-16 ***
animal_typeCAT             -0.18315    0.19396  -0.944 0.345017
animal_typeDOG             -0.23305    0.19251  -1.211 0.226055
animal_typeWILDLIFE        -0.38615    0.22984  -1.680 0.092945 .
month2                      0.14546    0.05533   2.629 0.008562 **
month3                     -0.25477    0.05692  -4.476 7.61e-06 ***
month4                     -0.28620    0.05657  -5.059 4.21e-07 ***
month5                     -0.11780    0.05181  -2.273 0.022996 *
month6                     -0.08372    0.04991  -1.677 0.093482 .
month7                     -0.14828    0.05051  -2.936 0.003327 **
month8                     -0.50955    0.05871  -8.679  < 2e-16 ***
month9                     -0.20499    0.05596  -3.663 0.000249 ***
month10                     0.05627    0.05167   1.089 0.276127
month11                    -0.05734    0.05436  -1.055 0.291483
month12                     0.16794    0.05157   3.257 0.001127 **
year2017                         NA         NA      NA       NA
intake_typeOWNER SURRENDER -0.77494    0.04076 -19.012  < 2e-16 ***
intake_typeSTRAY           -0.58951    0.03765 -15.658  < 2e-16 ***
chip_statusSCAN NO CHIP    -0.01469    0.02805  -0.524 0.600525
chip_statusUNABLE TO SCAN  -0.27589    0.06781  -4.068 4.73e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 10551.2  on 1449  degrees of freedom
Residual deviance:  9957.4  on 1431  degrees of freedom
AIC: 14017

Number of Fisher Scoring iterations: 6
```

So we can consider dropping the variable year and reconstruct the model with the remaining four variables as explanatory variables,just like the following model(2):

$$\log(\mathrm{E}(y_i)) = \beta_0 + \beta_{1i}\mathrm{animal\_type_i} + \beta_{2j}\mathrm{month_j} + \beta_{3p}\mathrm{intake\_type_p} + \beta_{4q}\mathrm{chip\_status_q}$$

where all the values of these variables remain unchanged and they are all group variables.

From the regression table below,we can see that the coefficients for each of the explanatory variables were essentially negative, and in terms of p-values, although some variables have p-values greater than 0.05 but given the correlated changes reflected in the box plots from the previous preliminary analysis, all explanatory variables included in the below regression model table are ultimately retained.

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

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-4.9857  -2.6038  -0.7551   0.8060  12.7653

Coefficients:
                           Estimate Std. Error z value Pr(>|z|)
(Intercept)                 2.75306    0.19690  13.982  < 2e-16 ***
animal_typeCAT             -0.18315    0.19396  -0.944 0.345017
animal_typeDOG             -0.23305    0.19251  -1.211 0.226055
animal_typeWILDLIFE        -0.38615    0.22984  -1.680 0.092945 .
month2                      0.14546    0.05533   2.629 0.008562 **
month3                     -0.25477    0.05692  -4.476 7.61e-06 ***
month4                     -0.28620    0.05657  -5.059 4.21e-07 ***
month5                     -0.11780    0.05181  -2.273 0.022996 *
month6                     -0.08372    0.04991  -1.677 0.093482 .
month7                     -0.14828    0.05051  -2.936 0.003327 **
month8                     -0.50955    0.05871  -8.679  < 2e-16 ***
month9                     -0.20499    0.05596  -3.663 0.000249 ***
month10                     0.05627    0.05167   1.089 0.276127
month11                    -0.05734    0.05436  -1.055 0.291483
month12                     0.16794    0.05157   3.257 0.001127 **
intake_typeOWNER SURRENDER -0.77494    0.04076 -19.012  < 2e-16 ***
intake_typeSTRAY           -0.58951    0.03765 -15.658  < 2e-16 ***
chip_statusSCAN NO CHIP    -0.01469    0.02805  -0.524 0.600525
chip_statusUNABLE TO SCAN  -0.27589    0.06781  -4.068 4.73e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 10551.2  on 1449  degrees of freedom
Residual deviance:  9957.4  on 1431  degrees of freedom
AIC: 14017

Number of Fisher Scoring iterations: 6
```

Finally,we transform the values in the model(2) into exponential formFinally,we transform the values in the model(2) into exponential form and plot the ratio of rates.In Figure 6, it can be seen that, from the

perspective of animal types, compared to birds, cats spent 0.83 times the number of days in the shelter, while dogs and other wild animals spent 0.79 and 0.68 times the number of days, respectively. In terms of the month the animal was admitted,compared to January, animals admitted in December spend 1.46 times the number of days in the shelter, which is the longest time, followed by animals admitted in February and October, while animals admitted in all other months spend less number of days in shelter than January, the smallest being animal admitted in August, which is 0.6 times compared to animal admitted in January.In terms of the reason for the animal being admitted,the confiscated animals stay longest time in shelter,compared to them, the number of days spend on shelter of animals with intake types of owner surrendered and stray are of 0.46 and 0.55 times respectively.For chip status,compared to the days spent in the shelter by animals with scan chips, animals without scan chips and those unable to be scanned were 0.99 and 0.76 times, respectively.

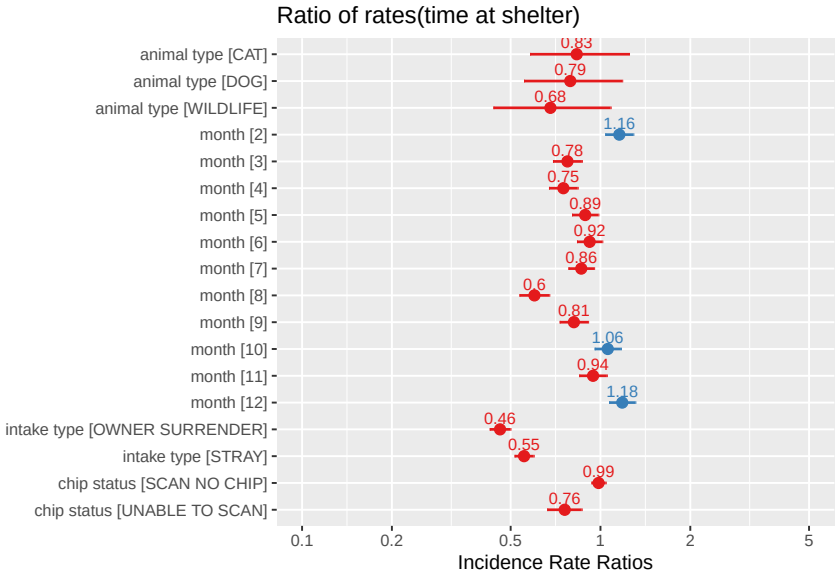

Figure 6: Figure 6:Ratio of rates(time at shelter)

Conclusion

In summary, in order to find What factors influence the days of animals spent at the shelter between being admitted and the final outcome (Time_at_Shelter),we constructed the model based on poisson GLM model with link function log to explore the relationship between the former and the explanatory variables(Animal_type,Month,Year,Intake_type,Chip_Status), and ultimately retains all variables except the year.Finally, by plotting the ratio of rates, we also explored the quantitative relationship between the variables within each type.