Animal_Rescue_in_NYC

Khine Hsu Wai

2024-02-03

R Markdown

Background_and_Introduction

This analysis is a part of Google data analytics certifcation. I received this dataset from NYC opendata link here

This data is made available by NYC department of parks and recreation. In this analysis, I will use this dataset to see the different distribution rescue and relocation of different types of animal species in NYC.

Table_of_Content

#Packages

Before we beigns, These are the required packages for this project on R, in additon to Tableau desktop and Tableau Public.

```
#library(readxl)
#install.packages("tidyverse")
#library(tidyverse)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
#install.packages("janitor")
#library(janitor)
#install.packages("timechange")
#library(timechange)
#install.packages("writexl")
#library(writexl)
```

Data Preparation

First we will start by importing the dataset to R. I directly uploaded the dataset from device to R. We want to see the summary statisc of Animal_resuce dataset first.

```
summary(Animal_resuce)
```

```
## Error in summary(Animal_resuce): object 'Animal_resuce' not found
```

```
Animal_rescue <- Animal_resuce
```

```
## Error in eval(expr, envir, enclos): object 'Animal_resuce' not found
```

Data Processing

I start this process by examining whether there are NA values first. Then, I checked to see how many type of Animals are there for the column Species Status and number of Animals.

First of all I will clean the data. First of all. Date and Time for initial call and response are in combined so I will spilt them. I also clean the age column as initally it has two or three different different variables in one row. I completed them under using pipe function. Finally, I deleted N/As from Animal Conditon, Species Status and # of Animals column. I also assigned different borough different code so that in the later part of the analysis, I can do logisitc regression testing. After this, I save the cleaned dataset as an excel file for my own record.

```
## Now Lets clean our data


# Date and Time for initial call and response are in combined so I will spilt them, I also clean the ag
Animal_rescue <- Animal_rescue %>%
  separate("Date and Time of initial call", c("date of initial call", "time of initial call"), sep = " 
  separate("Date and time of Ranger response", c("date of response", "time of response"), sep = " ") %>%
  mutate(cleaned_age = case_when(
    grepl("adult", Age, ignore.case = TRUE) ~ "adult",
    grepl("infant", Age, ignore.case = TRUE) ~ "infant",
    grepl("juvenile", Age, ignore.case = TRUE) ~ "juvenile",
    TRUE ~ NA_character_
  ))
```

```
## Error in separate(., "Date and time of Ranger response", c("date of response", : could not find funct
```

```
# changing the date to show only year
Animal_rescue <- Animal_rescue %>%
  mutate(Year_of_initial_call = year(as.Date(`date of initial call`, format="%Y-%m-%d")))
```

```
## Error in mutate(., Year_of_initial_call = year(as.Date(`date of initial call`, : object 'Animal_rescu
```

```
# Deleting N/As
Animal_rescue <- Animal_rescue %>%
```

```
  filter(`Animal Condition` != "N/A")
```

```
## Error in filter(., `Animal Condition` != "N/A"): object 'Animal_rescue' not found
```

```
Animal_rescue <- Animal_rescue %>%
  filter(`Species Status` != "N/A")
```

```
## Error in filter(., `Species Status` != "N/A"): object 'Animal_rescue' not found
```

```
Animal_rescue <- Animal_rescue %>%
  filter(`# of Animals` != "N/A")
```

```
## Error in filter(., `# of Animals` != "N/A"): object 'Animal_rescue' not found
```

```
# Assigning numbers to different borough to use later on
Animal_rescue <- Animal_rescue %>%
  mutate(Borough_Code = case_when(
    Borough == "Manhattan" ~ 1,
    Borough == "Brooklyn" ~ 2,
    Borough == "Queens" ~ 3,
    Borough == "Bronx" ~ 4,
    Borough == "Staten Island" ~ 5,
    TRUE ~ NA_integer_
  ))
```

```
## Error in mutate(., Borough_Code = case_when(Borough == "Manhattan" ~ 1, : object 'Animal_rescue' not
```

```
# Print the updated dataset
Animal_rescue
```

```
## Error in eval(expr, envir, enclos): object 'Animal_rescue' not found
```

```
# Save as Excel file
write_xlsx(Animal_rescue, "/Users/khinehsuwai/Desktop/Data_Analytics_Certificate.xlsx")
```

```
## Error in write_xlsx(Animal_rescue, "/Users/khinehsuwai/Desktop/Data_Analytics_Certificate.xlsx"): cou
```

Data_Analysis_and_Visualization

There will be two different type of analysis made in this sectioon, First the visualization created in R and uploaded from Tableau. The second part will be data analysis, where I test my hypothese of animals that are not from manhatttan get rescued later than the ones from other borough

R

Years_trend_in_animals_rescused

In this part, I used ggplot2, timechange(to convert the date column to be just year), dplyr and lubridate. I also changed the date to show only year as I want to see the trend as in groups of year. My inital hypothese for this analysis was that as in pandemic years, there was be more animal rescue due to many people not able to take care of their pets(domestic species) and thus seeing an increase in animal rescue in 2020 to 2022 for domestic species. However, surprisingly the only noticeable trend was at during 2019 there was an increase which peak at 2020 and goes doewn after 2020 for native species.

Plot1:Total_Number_of_Animals_by_Species_Status_and_Year

```
# Calculate Total_Count
Year_vs_Animal_Species<- Animal_rescue %>%
  group_by(Year_of_initial_call, `Species Status`) %>%
  summarise(Total_Count = sum(`# of Animals`, na.rm = TRUE)) %>%
  ungroup()
```

```
## Error in group_by(., Year_of_initial_call, `Species Status`): object 'Animal_rescue' not found
```

```
# Create a line graph
plot1<- ggplot(Year_vs_Animal_Species, aes(x = Year_of_initial_call, y = Total_Count, color = `Species S
  geom_line(stat = "identity") +
  geom_point() +
  labs(title = "Total Number of Animals by Species Status and Year",
       x = "Year",
       y = "Total Number of Animals",
       color = "Species Status") +
  theme_minimal()
```

```
## Error in ggplot(Year_vs_Animal_Species, aes(x = Year_of_initial_call, : could not find function "ggpl
```

```
print(plot1)
```

```
## Error in print(plot1): object 'plot1' not found
```

```
# Save the plot as a PNG file
ggsave("plot1.png", plot1, width = 10, height = 6, units = "in")
```

```
## Error in ggsave("plot1.png", plot1, width = 10, height = 6, units = "in"): could not find function "g
```

```
# Print a message indicating that the plot has been saved
print("plot1.png")
```

```
## [1] "plot1.png"
```

The hypothese was further proven to be not ture by native species on top of the most animal rescued in 2020 to 2022

```
filtered_data <- Year_vs_Animal_Species %>%
  filter(Year_of_initial_call>= 2020 & Year_of_initial_call <= 2022)
```

```
## Error in filter(., Year_of_initial_call >= 2020 & Year_of_initial_call <= : object 'Year_vs_Animal_Sp
```

```
# Group by species status and calculate the total rescue count for each species status
species_summary <- filtered_data %>%
  group_by(`Species Status`) %>%
  summarise(Total_Rescue = sum(`Total_Count`, na.rm = TRUE))
```

```
## Error in group_by(., `Species Status`): object 'filtered_data' not found
```

```
# Arrange in descending order to get the top 5
top_species <- species_summary %>%
  arrange(desc(Total_Rescue)) %>%
  head(5)
```

```
## Error in arrange(., desc(Total_Rescue)): object 'species_summary' not found
```

```
# Print the top 5 species
top_species
```

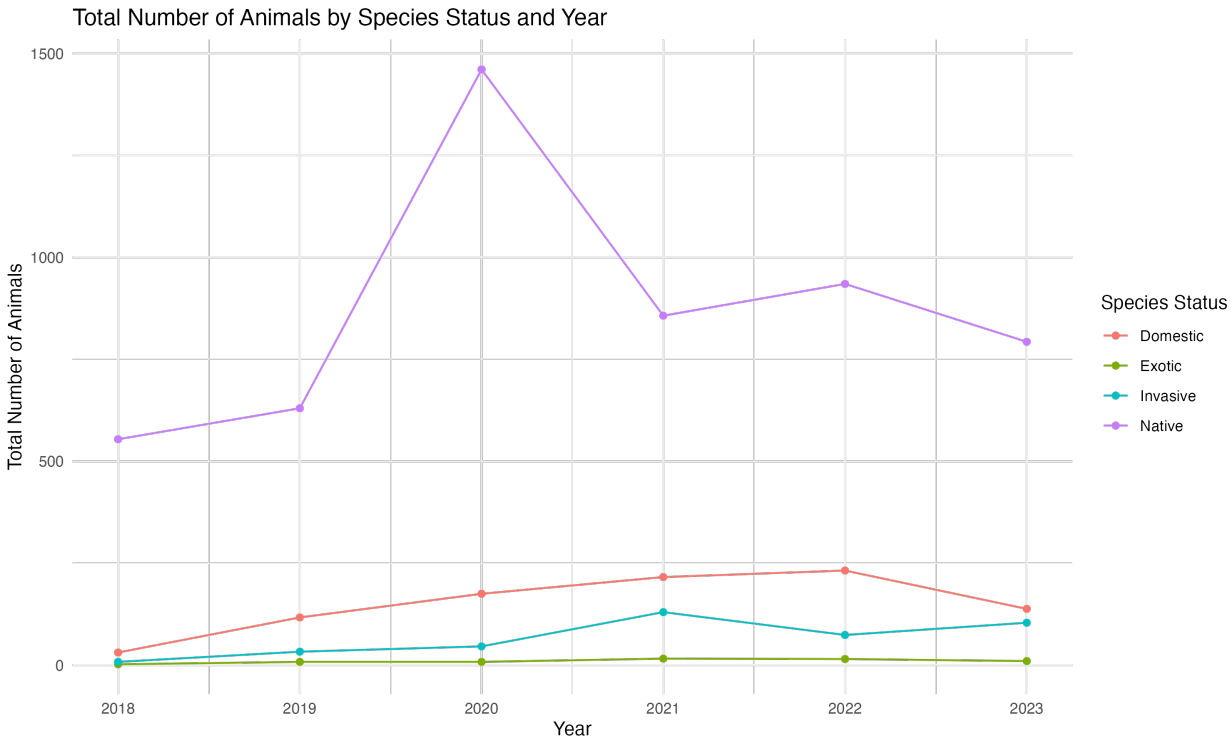


Figure 1: Plot

```
## Error in eval(expr, envir, enclos): object 'top_species' not found
```

Plot2:The__most__rescured__Animal__Species

In my second Analysis, I want to see what is the the details insight on each animal species number that were rescued.

```
library(ggplot2)

plot2<-ggplot(Animal_rescue, aes(x = `Species Status`, fill = `Species Status`, weight = `# of Animals`)
  geom_bar() +
  geom_text(
    aes(label = after_stat(count)),
    stat = "count",
    position = position_stack(vjust = 0.5),
    size = 3
  ) +
  labs(title = "Number of Animals by Species Status",
       x = "Species Status",
       y = "Total Number of Animals") +
  theme_minimal()
```

```
## Error in ggplot(Animal_rescue, aes(x = `Species Status`, fill = `Species Status`, : object 'Animal_re
```

```
ggsave("plot2.png", plot2, width = 10, height = 6, units = "in")
```

```
## Error in plot_theme(plot): object 'plot2' not found
```

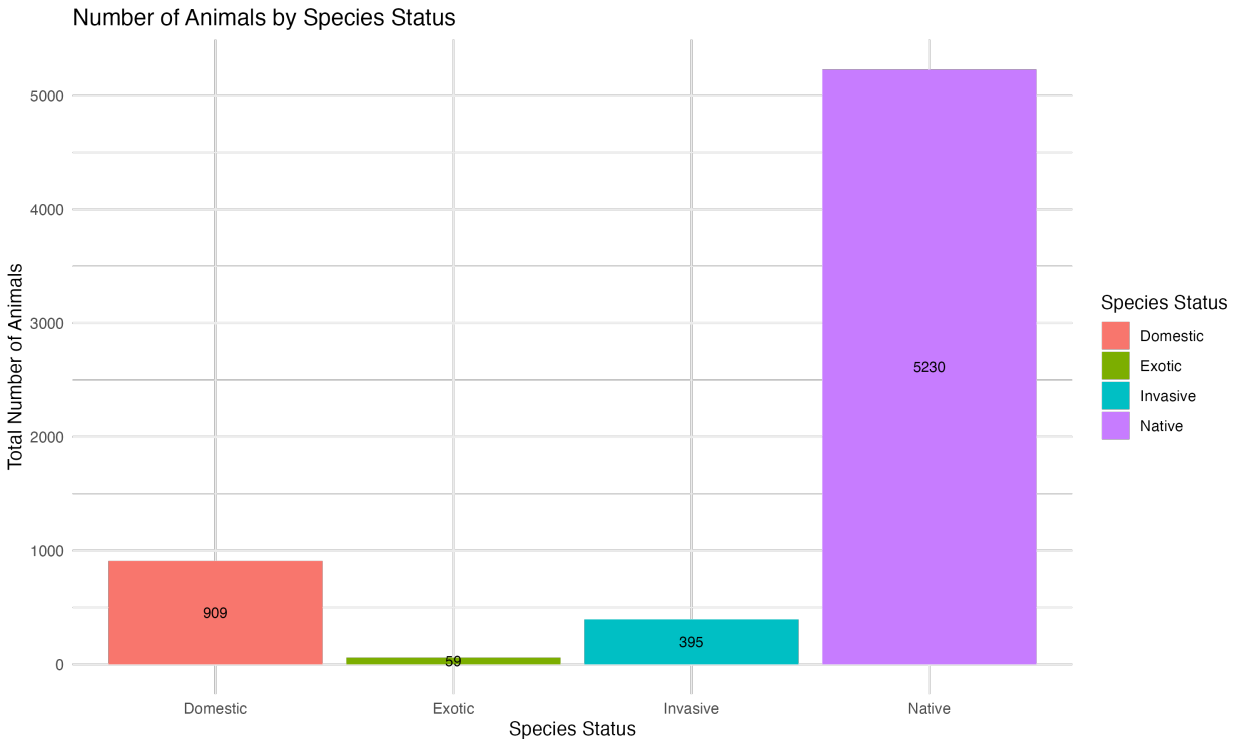
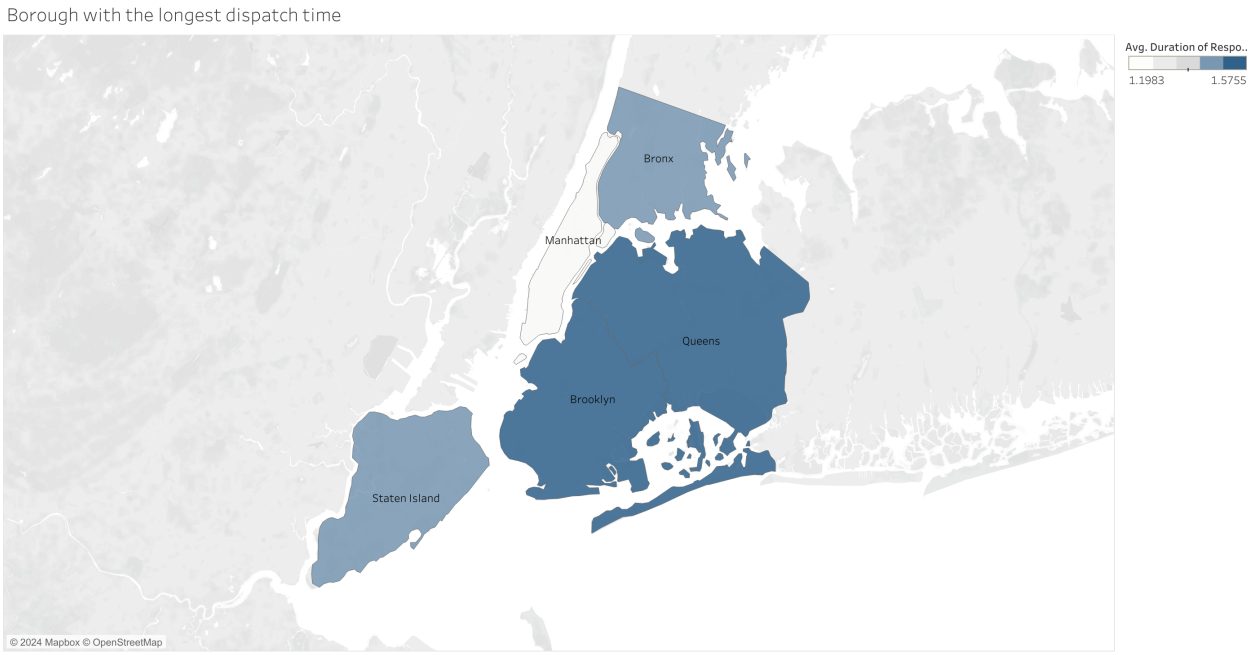


Figure 2: Plot

Tableau Analysis

Plot1:Boroughs_with_the_longest_dispatch_time

My first analysis is based on the hyphothese that manhattan would be the fastest neighborhood in terms of dispatch time and I proved that it is right by my visualization in tableau with Manhattan being the fastest and Queens being the latest.



Plot2:Boroughs_in_Ranks_of_Animal_Conditions_when_being_rescued

My second analysis is developed after the first hyphothese that since manhattan being the fastest neighborhood when it comes to animal rescused, that it would also have the highest amount of healthy or least amount of DOA(dead on arrival) animals compared to other borough. However, from my analysis, it would seems so that my hyphothese would be debunked. With Manhattan having the largest number of DOA and having third lowest number of healthy animals being rescued.

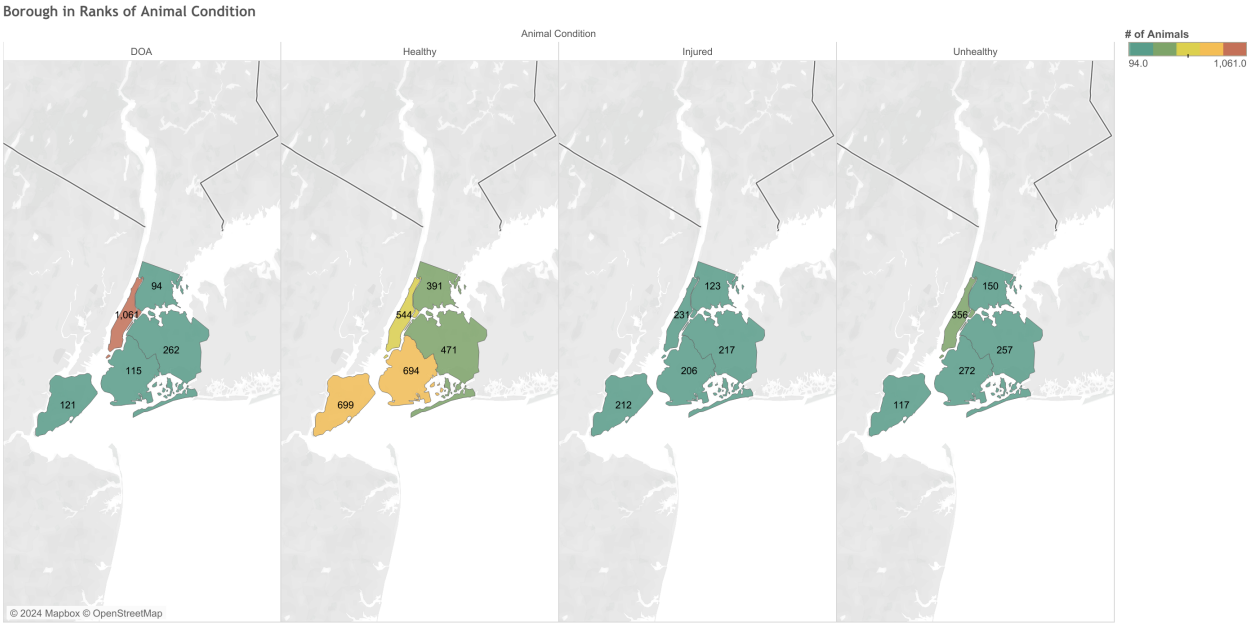


Figure 3: Animal_condition_on_different_age.png

Plot3:Animal Condition vs Different Age Group

My third analysis is created to see if different age groups have effect on animal's condition upon rescued. To do that I group three different age group(adult,Juvenille and infant) and then grouped conditions(DOA,healthy,unhealthy and injured). After that, I compared it with number of animals. Upon my surprise, there is no immediate connection between two. However, I rather find other interesting fact that adults are the one that get most rescued out of all age groups.

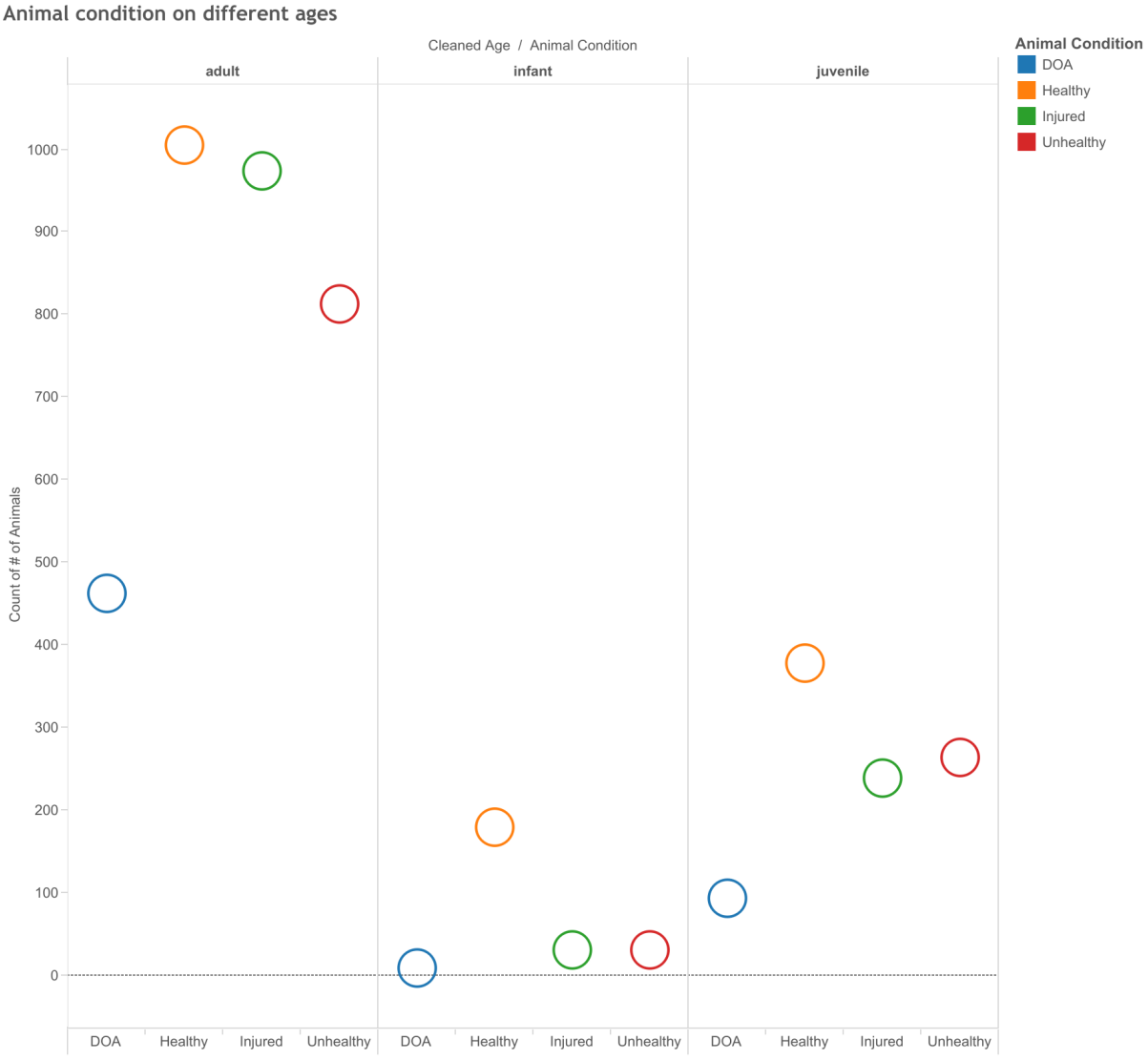


Figure 4: Animal_condition_on_different_age

Conclusion

Business_Suggestion

This whole assignment not only represents on status upon Animal rescued but also on the fact that the different status these animals faced. For example, although Manhattan has the fastest dispatch rate, it also has the highest dead on arrival rate. As Manhattan is one of the New York City's popular place where the image represents a very important role. As well as it is one of the most busiest cities in the world thus many

traffic collisions. Perhaps, these are the driving factors behind many animals getting dispatched fast but ended up passing away.

Case_study

This is my first time finishing the whole dataset outside of school to finish on my own for my passion. As someone who loves animal,I had alot of fun working oh this dataset. Please let me know any feedbacks or comments for imporvements.

Resources

1. NYC Open Data link here by Department of Parks and Recreation (DPR).
2. Tableau Desktop
3. Tableau Public
4. R Programming
5. Excel