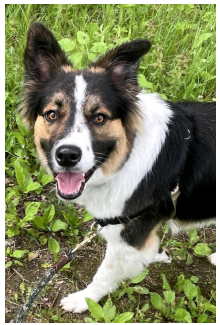
Waupaca Humane Spay and Neuter Analysis

Presented By:
Karganleh Borh, Alex Mitchell, Carolyn Shaffer, and Jeff Sischo

Introduction

The purpose of this project is to analyze data from the Waupaca Humane society. The analysis will support a grant request for Spaying and Neutering free roaming cats and pets of low income families.



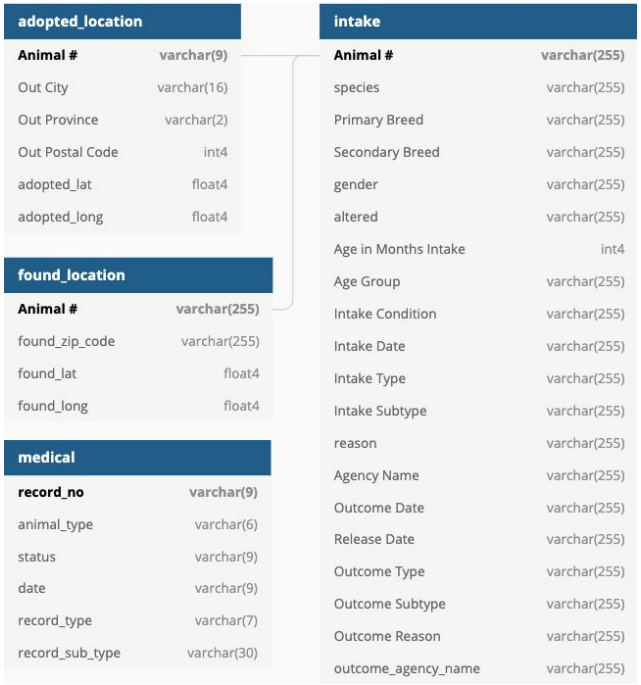
Overview

- Data cleansing and storing
- General information and trends
- Spay and neuter analysis
- Spay and neuter predictions

Every Row is a Wet Nose

Data

- Provided with 6 Excel files including medical and intake data over a time span of 3 years
- Performed preliminary cleaning and data organization in Excel
- Performed additional data cleansing with Python to ensure data met our model and dashboards needs
- Designed relational base
- Created database to store data hosted on Amazon RDS
- Connected database to machine learning model



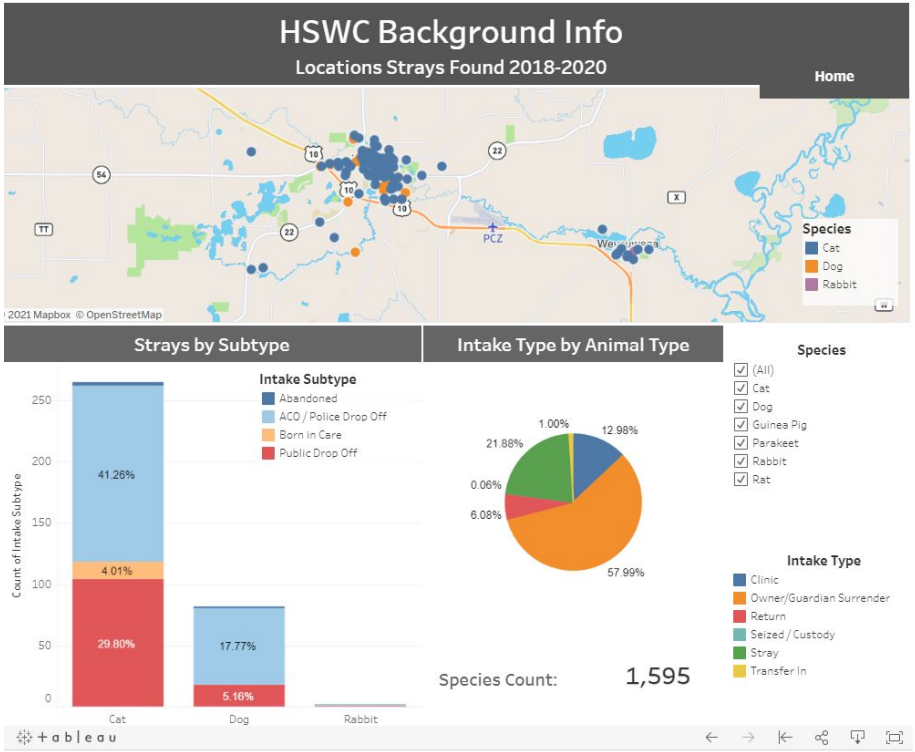
Waupaca Humane Serves the Community's Animal Population

HSWC Dashboards

- Background Stats
- Spay & Neuter Info
- Projected Spay & Neuter Needs

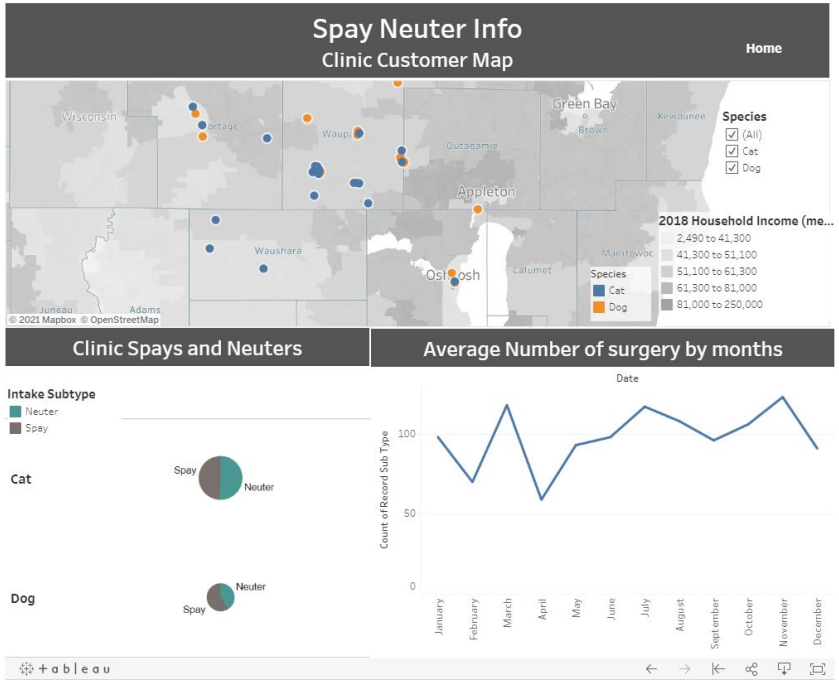
HSWC Background Info

- Map of areas served by the Waupaca Humane Society with locations stray animals were found (2018-2020)
- Percentage of intake types by species
- Comparison of Stray intake by subtype for cats vs dogs
- Overall percentages of intake types for all animals



Spay and Neuter Dashboards

- Map of Waupaca Humane Society Clinic patient with income level overly
- Percentage of surgeries performed by species
- Surgeries performed overtime



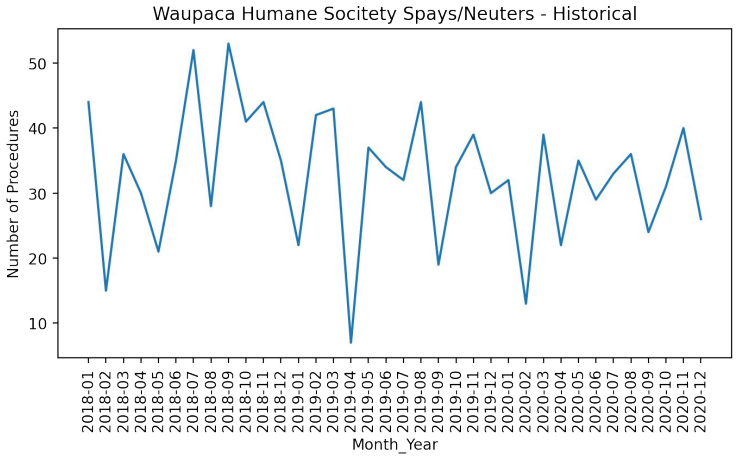
Dashboard Demo

HSWC Dashboard

Waupaca Humane seeks funding to continue Spay and Neuter Services to Low Income families and Free Roaming Cats

Choice of ML model - ARIMA

- ARIMA (AutoRegressive Integrated Moving Average) Model.
  - Plot of cleaned data for initial visualization.



Choice of ML model - ARIMA

- **Limitations/Benefits**
  - **Outliers can have a large effect on the forecast and they can't be filtered out easily.**
  - **Univariate - only one series being looked at. Can have a large time series and it is easy to handle.**
  - **Simple enough to not overfit.**

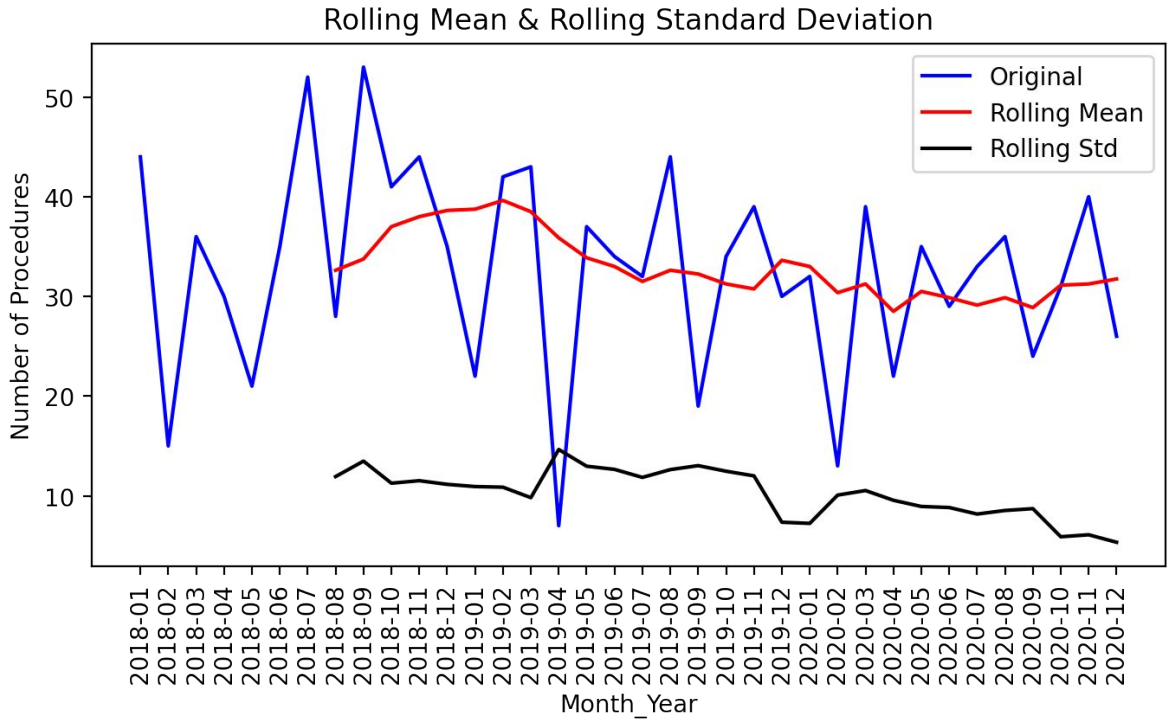
Data Manipulation - ARIMA

- Features and Engineering
  - Main Features: Date and record_sub_type (spay/neuter)
  - Engineering that needed be completed:
    - Extracting M/Y from M/D/Y format
    - Count of the spay/neuter procedures performed
- Split of Test/Train
  - Data series was split into test/train groups of 75% Train and 25% Test:
    - X = counts.values
    - size = int(len(X) * 0.75)
    - train, test = X[0:size], X[size:len(X)]

Stationary Data Check

Rolling Statistics method

Used 8 period roll



Stationary Data Check

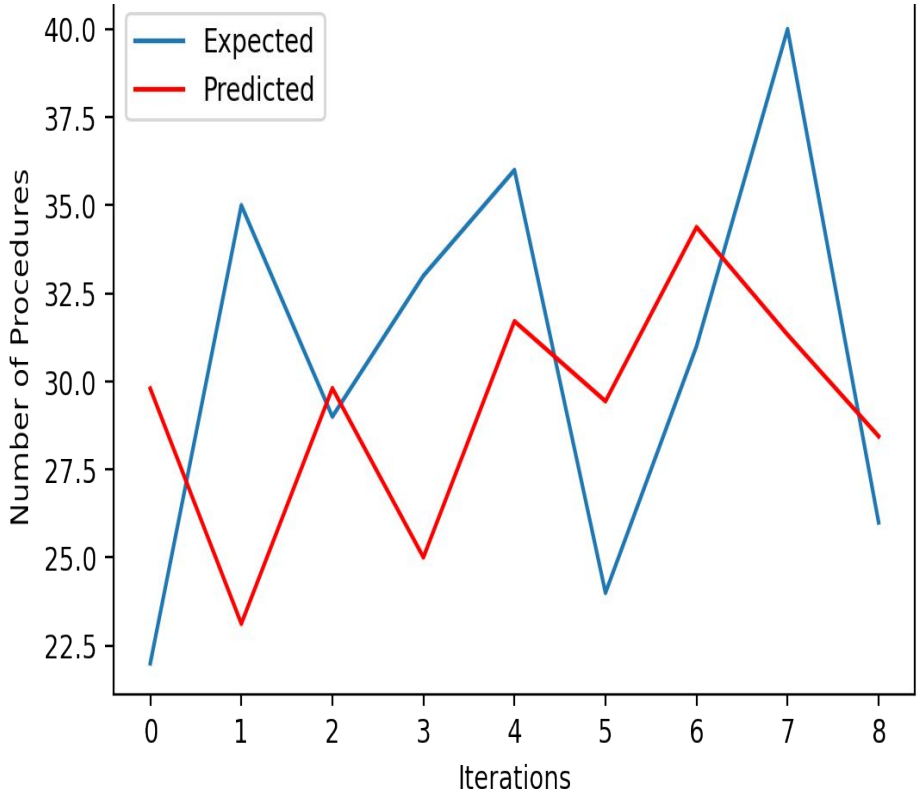
Augmented Dickey-Fuller (ADF) statistical check

- Desired results:
  - Low p-value
  - 1%, 5% and 10% critical values CIs are close to the ADF.

```
ADF Statistic: -4.523442265864301
p-value: 0.00017832531117933361
Critical Values:
        1%: -3.639224104416853
        5%: -2.9512301791166293
        10%: -2.614446989619377
```

Validation of Model

```
predicted=29.794556, expected=22.000000
predicted=23.122617, expected=35.000000
predicted=29.805987, expected=29.000000
predicted=25.007105, expected=33.000000
predicted=31.708055, expected=36.000000
predicted=29.439180, expected=24.000000
predicted=34.379528, expected=31.000000
predicted=31.325659, expected=40.000000
predicted=28.443528, expected=26.000000
Test RMSE: 6.725
```



Final ARIMA Projection

HSWC Dashboard

Conclusion

Outcomes:

- Continued need to Waupaca to perform these surgeries
  - Waupaca will perform about 30 surgeries a month for the next 3 years
  - Clear need for spay and neuter services for cats
  - Data Supports ability to perform surgeries

Recommendations:

- Additional analysis on seasonality

What would we do differently?

- More Data
- A relationship between intake tables and medical tables