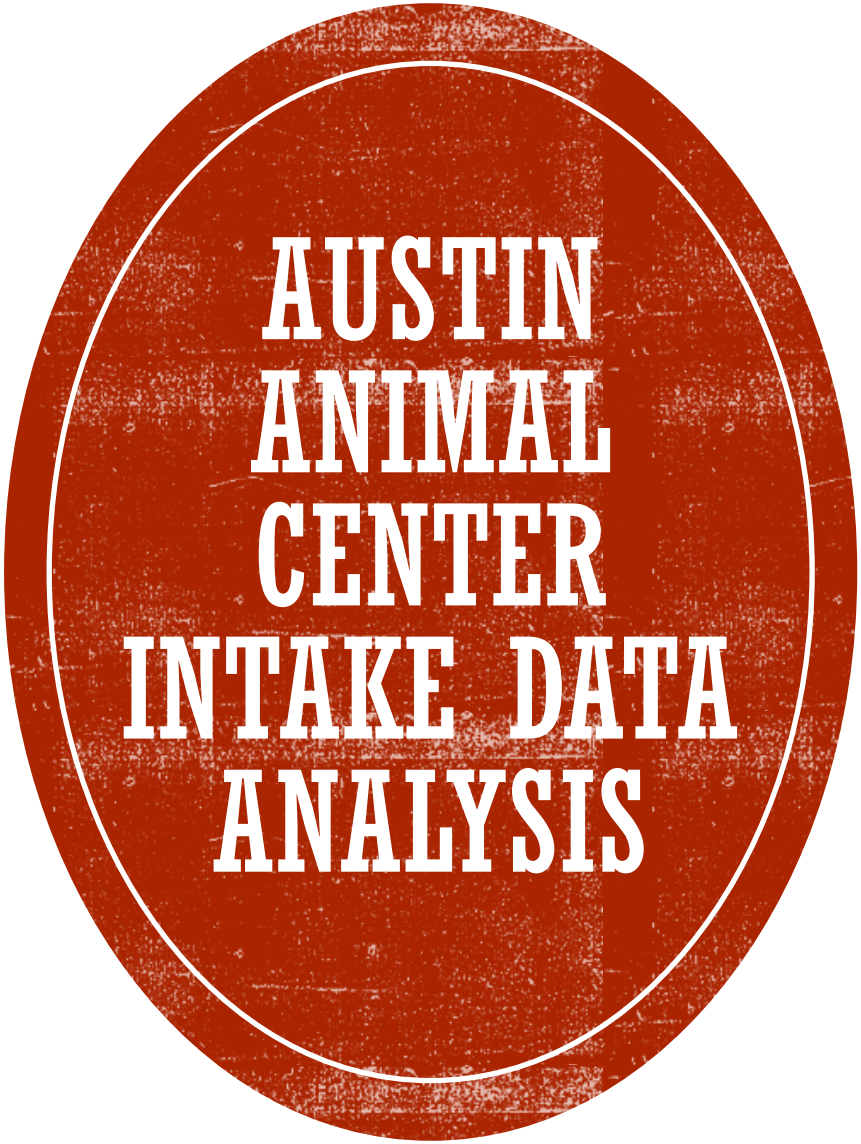


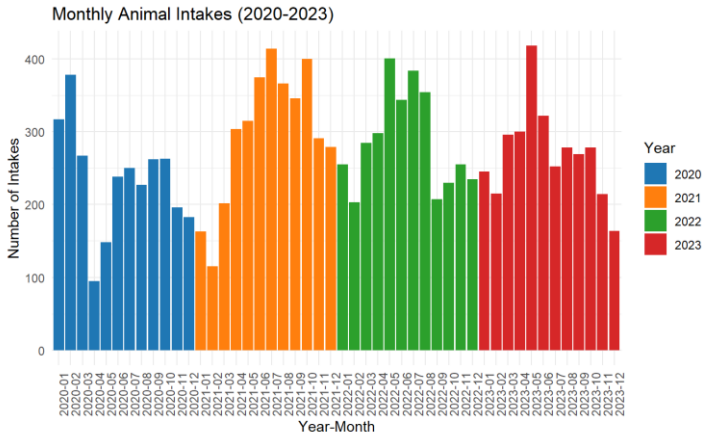
Exploring Trends, Breeds, and Shelter Insights

By Hannah Marques

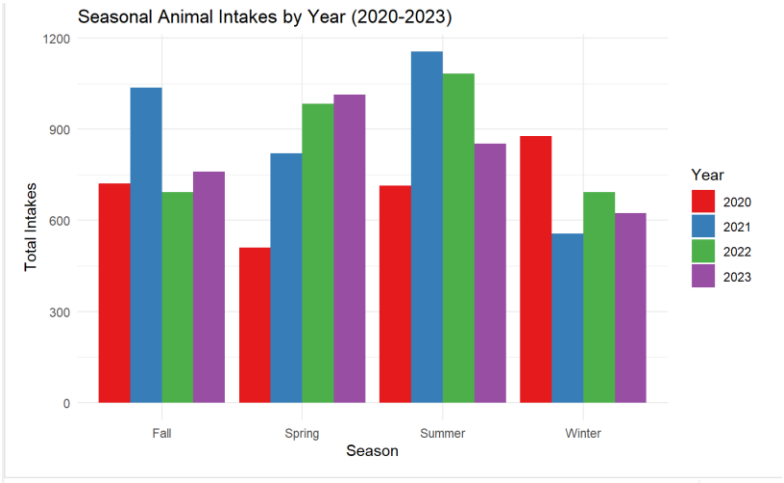
INTRODUCTION

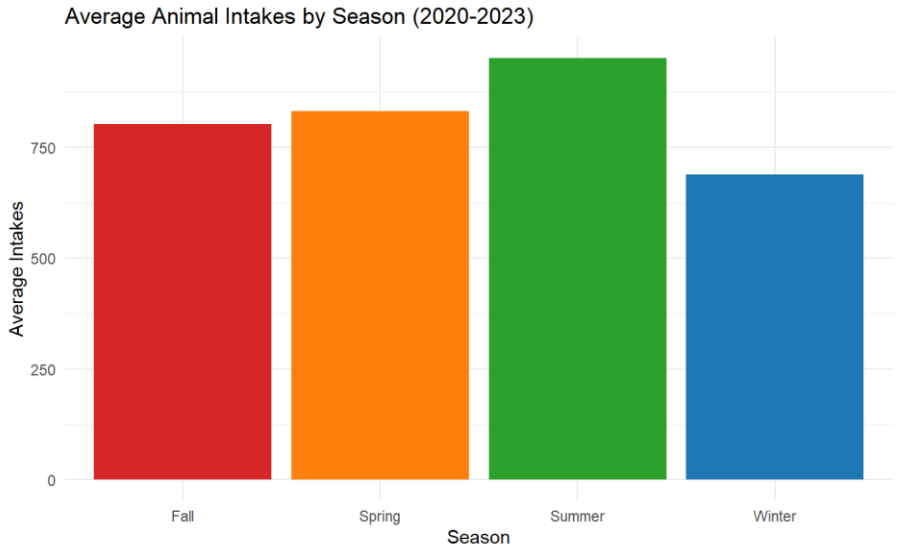
- **Purpose:** Analyze animal intake trends at Austin Animal Center (2020-2023).
- **Focus Areas:** Seasonal trends, breed-specific patterns, and shelter capacity.
- **Research Questions:**
  - What seasonal patterns emerge?
  - How do intake types vary across breeds?
  - Are there high-intake months that challenge shelter capacity?
  - What is the maximum capacity of the Austin Animal Center Shelter? Was there any month in this analysis where the shelters reached capacity?

METHODS



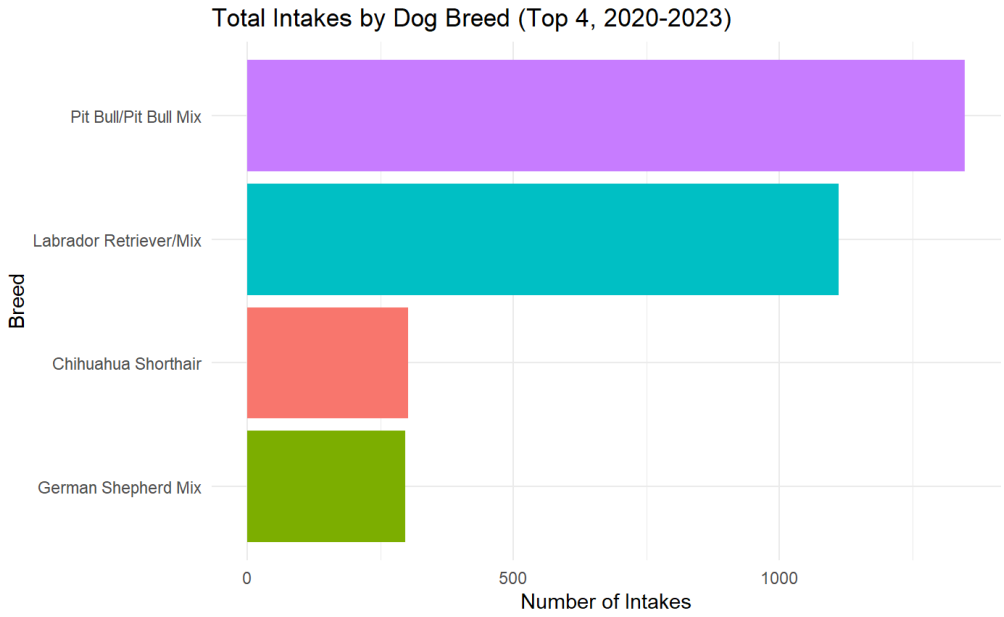
- **Data Cleaning:** Filtered for 2020-2023 data, grouped breeds and created seasonal variables.
- **Variables:** Combined breed categories (e.g., Pit Bull types) and added binary variables for models.
- **Tools:** Used R for data visualization, statistical summaries, and logistic regression modeling.

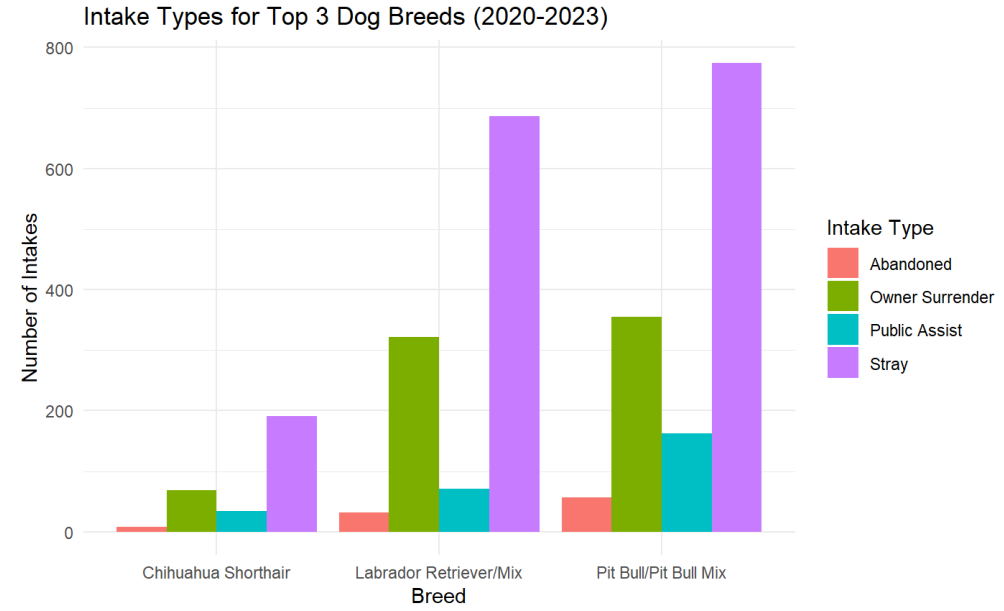




EXPLORATORY DATA ANALYSIS (SEASONAL TRENDS)

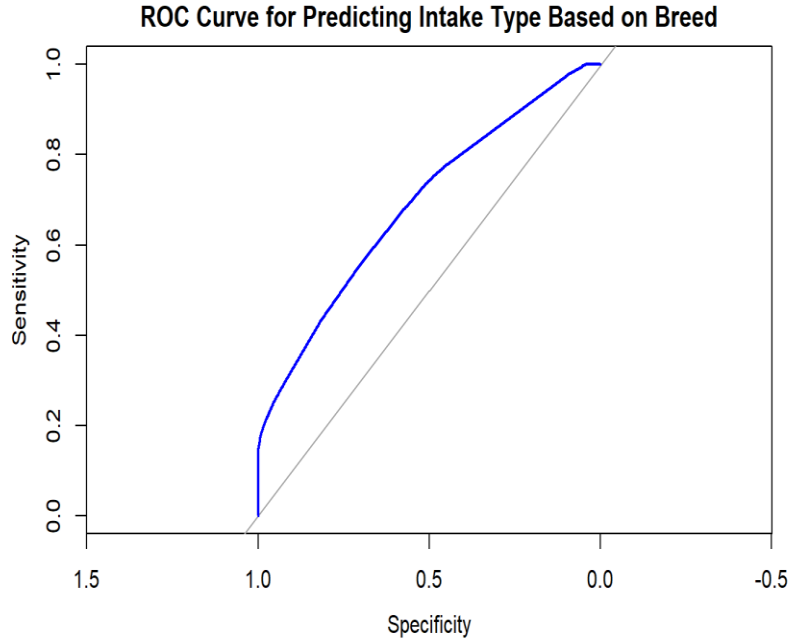
- **Highest** average intakes observed in **Summer** (948.5 intakes).
- **Winter** has the **lowest** average intakes (706 intakes).
- Seasonal spikes likely influenced by **breeding** and **abandonment patterns.**





EDA RESULTS (BREED ANALYSIS)

- **Pit Bull/Mix** and **Labrador Retriever/Mix** types and mixes have disproportionately high intakes.
- **Public Assist** and **Owner Surrender** are common for breeds with **stigmas.**
- **Rare breeds** show **limited data**, challenging predictive analysis.




```r
# Convert "Stray" to binary outcome
data$IntakeTypeBinary <- ifelse(data$`Intake Type` == "Stray", 1, 0)

# Use the existing "Season" column to create a "IsSummer" predictor
data$IsSummer <- ifelse(data$Season == "Summer", 1, 0)

# Ensure categorical variables are treated as factors
data$Breed <- as.factor(data$Breed)

# Fit Logistic Regression Model with predictors: Breed and IsSummer
logistic_model <- glm(IntakeTypeBinary ~ Breed + IsSummer, family = binomial(), data = data)

# Display Key Coefficients and Model Metrics
summary_log <- summary(logistic_model)
cat("Coefficients (Top Predictors):\n")
print(head(summary_log$coefficients, 10))  # Show only top 10 predictors
cat("\nModel Deviance:\n")
cat("Null Deviance:", summary_log$null.deviance, "\n")
cat("Residual Deviance:", summary_log$deviance, "\n")
cat("AIC:", AIC(logistic_model), "\n")

# Generate Predicted Probabilities
data$PredictedProb <- predict(logistic_model, type = "response")

# Calculate ROC and AUC
roc_curve <- roc(data$IntakeTypeBinary, data$PredictedProb)
auc_value <- auc(roc_curve)
cat("\nArea Under the Curve (AUC):", auc_value, "\n")

# Plot ROC Curve
plot(roc_curve, col = "blue", main = "ROC Curve for Predicting Likelihood of Stray Intake in Summer")
```

MODEL INSIGHTS

- **Logistic Regression:** Predicting intake type based on breed.
- **Key Metrics:**
  - AUC: **0.6997** (Moderate predictive power).
- **Challenges:** Data imbalance for rare breeds impacts model performance.

```r
# Prepare data for cross-validation
set.seed(123)  # For reproducibility
folds <- createFolds(data$IntakeTypeBinary, k = 5, list = TRUE, returnTrain = TRUE)

# Define a function to train and evaluate the model on each fold
cv_results <- lapply(folds, function(train_index) {
  # Split data into training and test sets
  train_data <- data[train_index, ]
  test_data <- data[-train_index, ]

  # Ensure test data has the same levels as training data
  train_data$Breed <- droplevels(train_data$Breed)  # Drop unused levels in training data
  test_data$Breed <- factor(test_data$Breed, levels = levels(train_data$Breed))

  # Remove rows with unseen levels
  test_data <- test_data[complete.cases(test_data), ]

  # Check if test_data is empty after filtering
  if (nrow(test_data) == 0) {
    return(NA)  # Return NA if no valid test data exists
  }

  # Fit logistic regression model
  logistic_model <- glm(IntakeTypeBinary ~ Breed + IsSummer, family = binomial(), data = train_data)

  # Predict on test set
  test_data$PredictedProb <- predict(logistic_model, newdata = test_data, type = "response")

  # Calculate AUC for the fold
  roc_curve <- roc(test_data$IntakeTypeBinary, test_data$PredictedProb)
  auc_value <- auc(roc_curve)

  return(auc_value)
})

# Calculate average AUC and consistency
valid_auc_values <- unlist(cv_results)
valid_auc_values <- valid_auc_values[!is.na(valid_auc_values)]  # Remove NA values if any
average_auc <- mean(valid_auc_values)
std_dev_auc <- sd(valid_auc_values)

# Print results
cat("5-Fold Cross-Validation Results:\n")
cat("Average AUC:", average_auc, "\n")
cat("Standard Deviation of AUC:", std_dev_auc, "\n")
```

CROSS-VALIDATION SUMMARY

- **Objective:**
  Evaluate the model's ability to predict stray intakes using 5-fold cross-validation.
- **Results:**
- **Average AUC:** 0.5911 (slightly better than random guessing).
- **Standard Deviation of AUC:** 0.0095 (consistent performance across folds).
- **Conclusion:**
  The model is stable but has limited predictive power with breed and seasonality as predictors. But it is sufficient to see increasing trends due to weather. Additional factors like intake condition or age could improve accuracy.

DISCUSSION

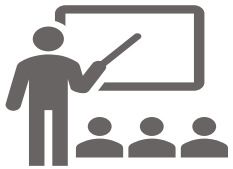
Key Findings:

Seasonal spikes in **Summer** and Beginning of **Fall.**

Overrepresentation of certain breeds like **Pit Bulls.**

Covid-19 Pandemic Affected the Intake rates, which picked up quickly after.

Intake numbers are constantly higher than shelter **capacity** for most months.



Next Steps:

Address **stray** population trends.

Educate on breed stigma and promote **pet friendly housing policies.**

Address shelter capacity issues and take measures to prevent it.

REFERENCES

- Human Animal Support Services (2023). The Problem with Pit Bulls. www.humananimalsupportservices.org/blog/the-problem-with-pit-bulls/
- Let Love Live (2023). Why Pit Bull Mixes Dominate Southeastern Shelters. www.letlovelive.org/uncategorized/why-pit-bull-mixes-dominate-southeastern-shelters/
- Austin Animal Center (2023). Shelter Capacity. www.austintexas.gov/page/shelter-capacity
- City of Austin. (n.d.). *Austin Animal Center Intakes*. Retrieved from data.austintexas.gov/Health-and-Community-Services/Austin-Animal-Center-Intakes/wter-evkm