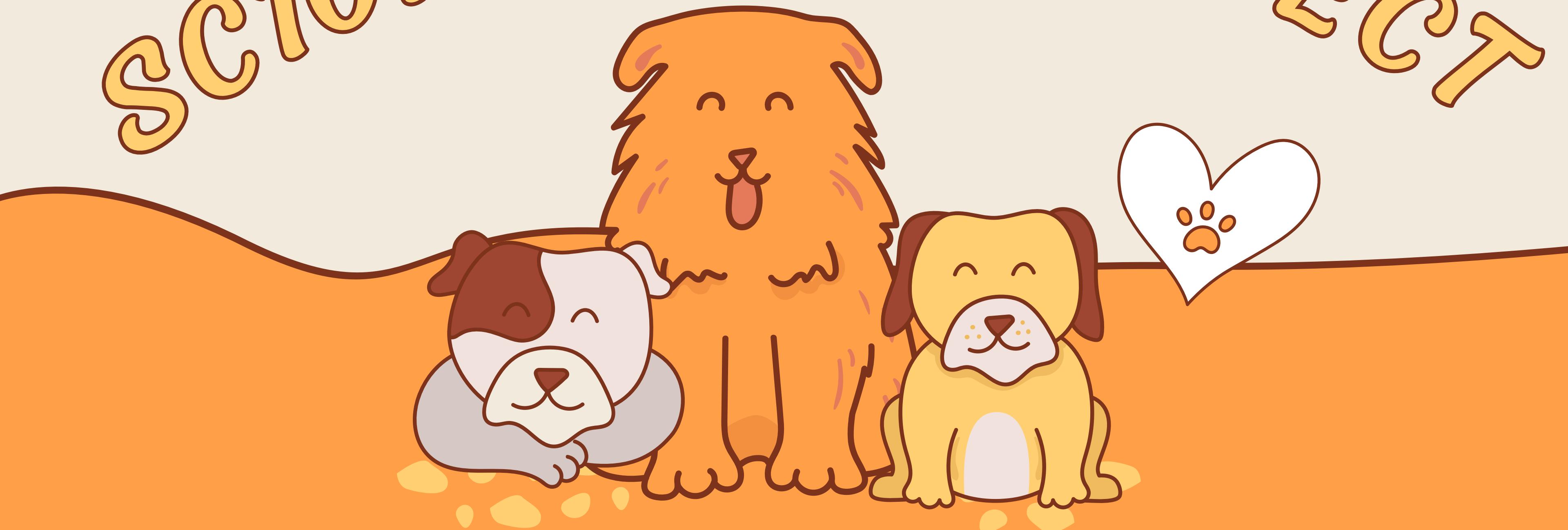


SC1015 MINI PROJECT



OPTIMIZING SHELTER MANAGEMENT AND BOOSTING ADOPTION RATES

FCCA TEAM8

JIAO YUNQI & KWOK XIN TZI ELLIS & JEANNIE WONG YI LIN

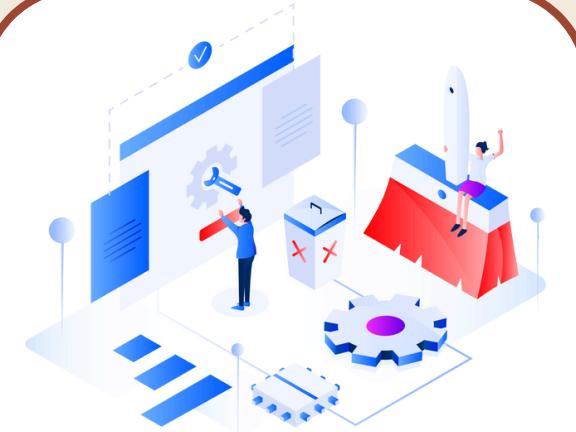
PRES

ENTATION

ROADMAP



Motivation



Data Cleaning



Exploratory Data Analysis



Machine Learning



Data-driven Insights



Conclusion



MOTIVATION



PROBLEM DEFINITION



HOW DO WE OPTIMIZE THE
MANAGEMENT OF ANIMAL SHELTERS AND
INCREASE ADOPTION RATES?

WHERE THE DATA FROM

data.austintexas.gov
the official City of Austin open data portal

Austin Animal Center Intakes

Health and Community Services

Animal Center Intakes from Oct, 1st 2013 to present. Intakes represent the status of animals as they arrive at the Animal Center. All animals receive a unique Animal ID during intake. Annually over 90% of animals entering the center, are adopted, transferred to rescue or returned to their owners.



Data Used:

2024.3.27

Last Updated
April 20, 2024

Data Provided By
City of Austin, Texas -
data.austintexas.gov

Austin Animal Center Outcomes

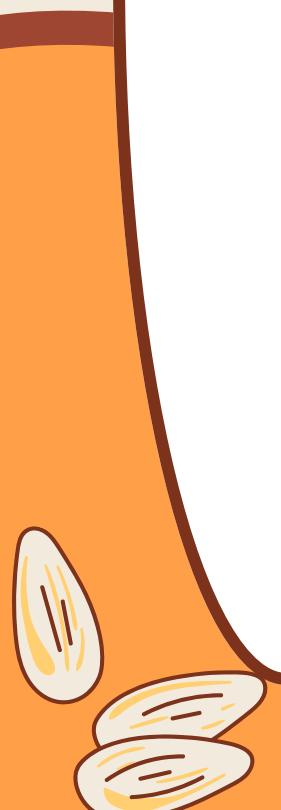
Health and Community Services

Animal Center Outcomes from Oct, 1st 2013 to present. Outcomes represent the status of animals as they leave the Animal Center. All animals receive a unique Animal ID during intake. Annually over 90% of animals entering the center, are adopted, transferred to rescue or returned to their owners. The Outcomes data set reflects that Austin, TX. i...

[Read more ▾](#)

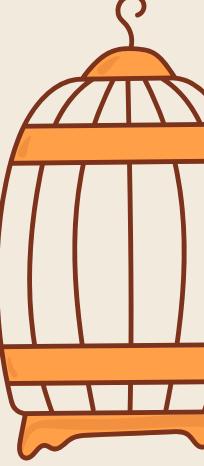
Last Updated
April 20, 2024

Data Provided By
City of Austin, Texas -
data.austintexas.gov





DATA CLEANING



DATA CLEANING

STEP 01

Sort and Differentiate
Animal ID



STEP 02

Sort Data and
Handle Duplicate Records



STEP 03

Merge Intake and
Outcome Dataset



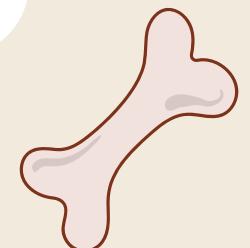
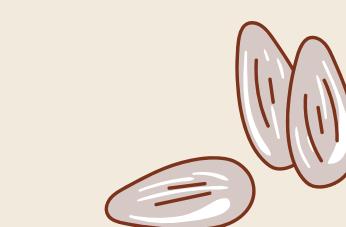
STEP 05

Feature Extraction

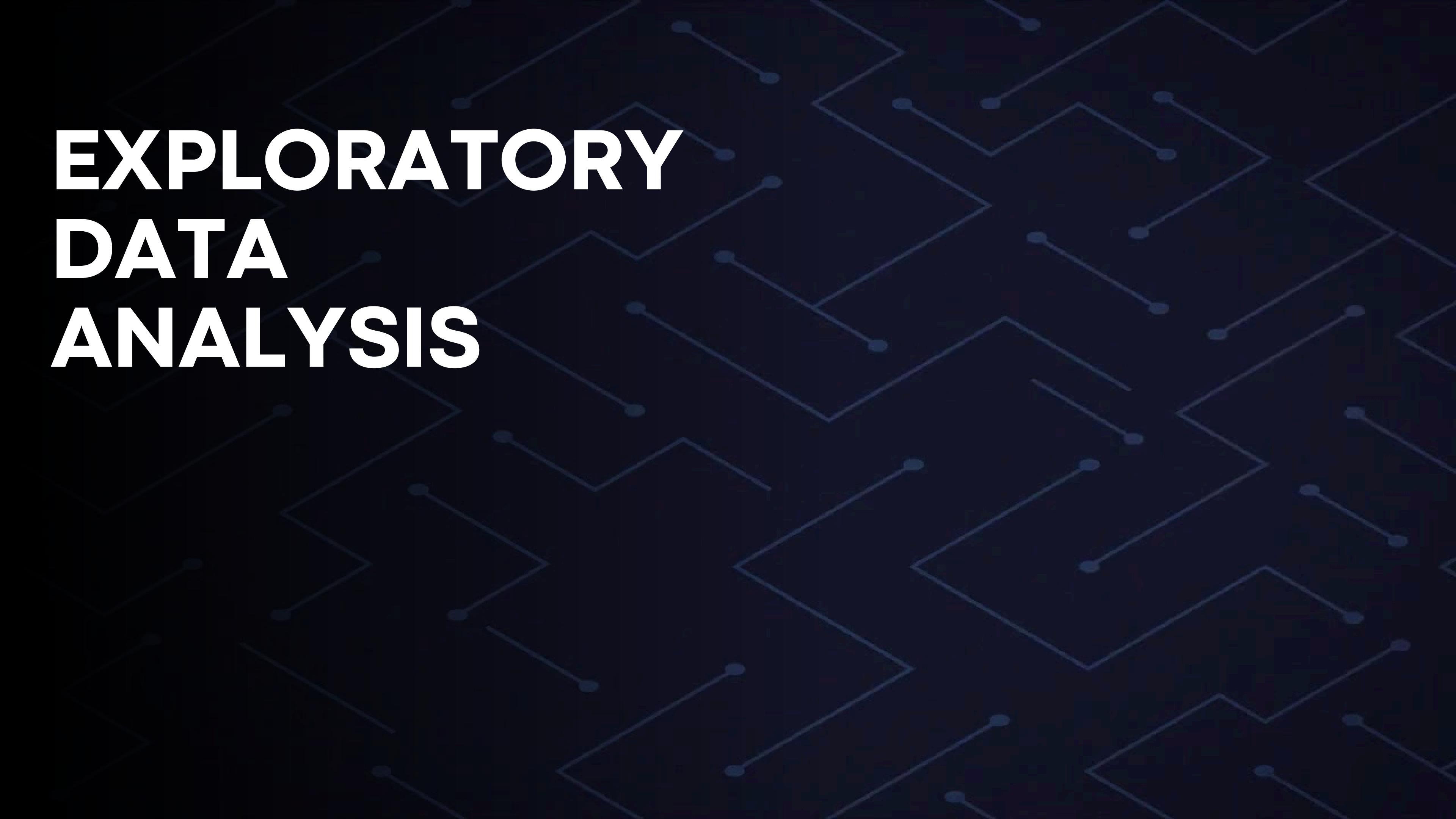


FINISH

Export Clean CSV



EXPLORATORY DATA ANALYSIS

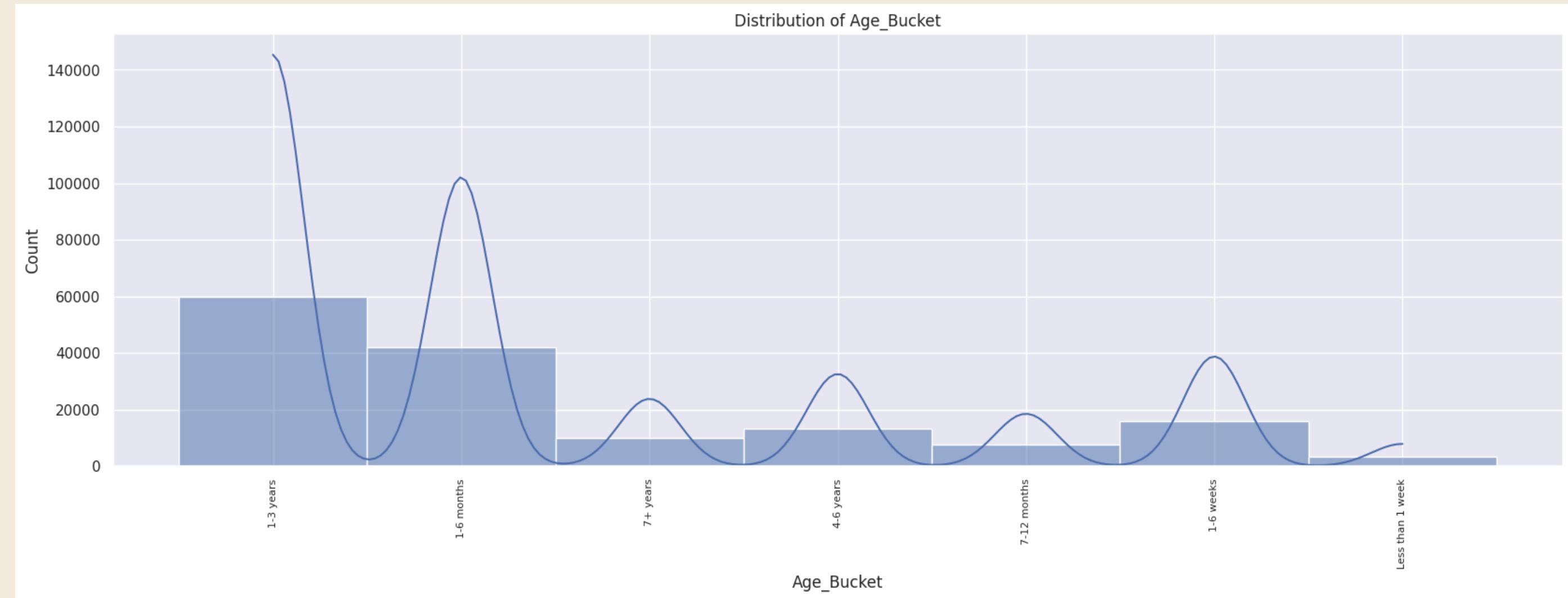


NUMERICAL DATA



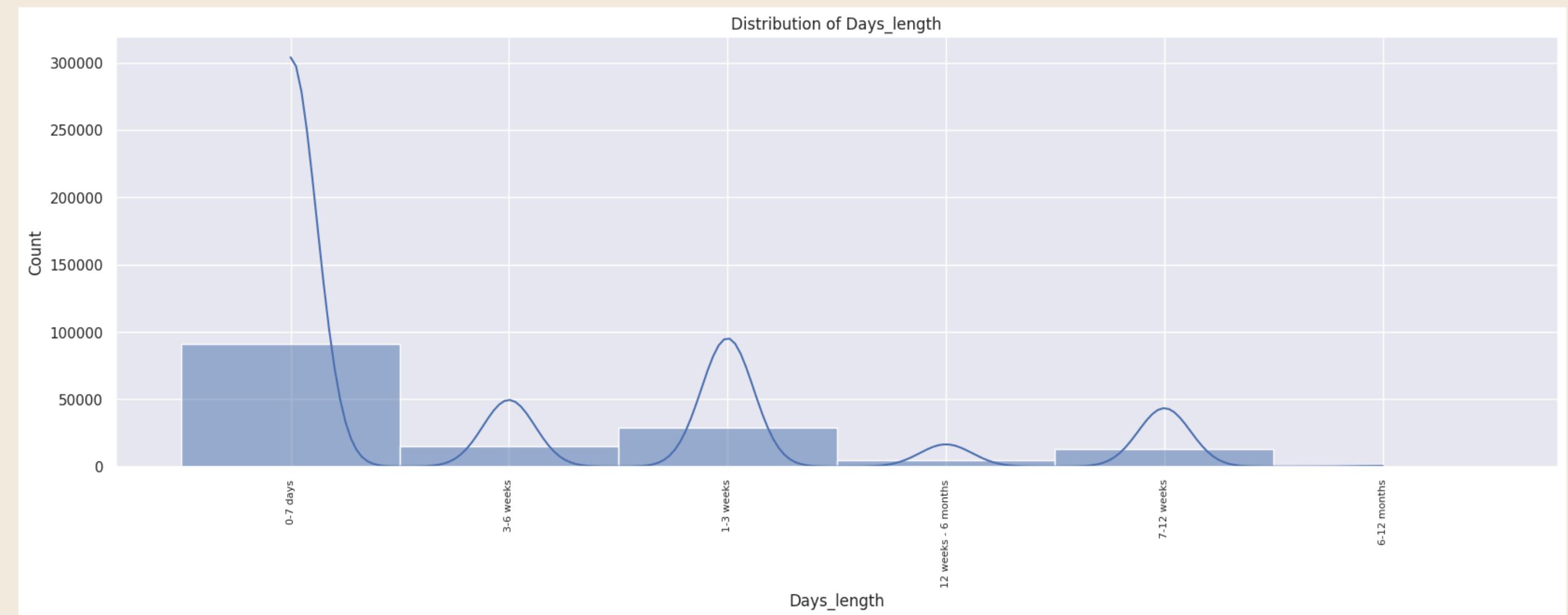
DISTRIBUTIONS OF NUMERICAL FEATURES

Distribution of Age



DISTRIBUTIONS OF NUMERICAL FEATURES

Distribution of Day Length of stay



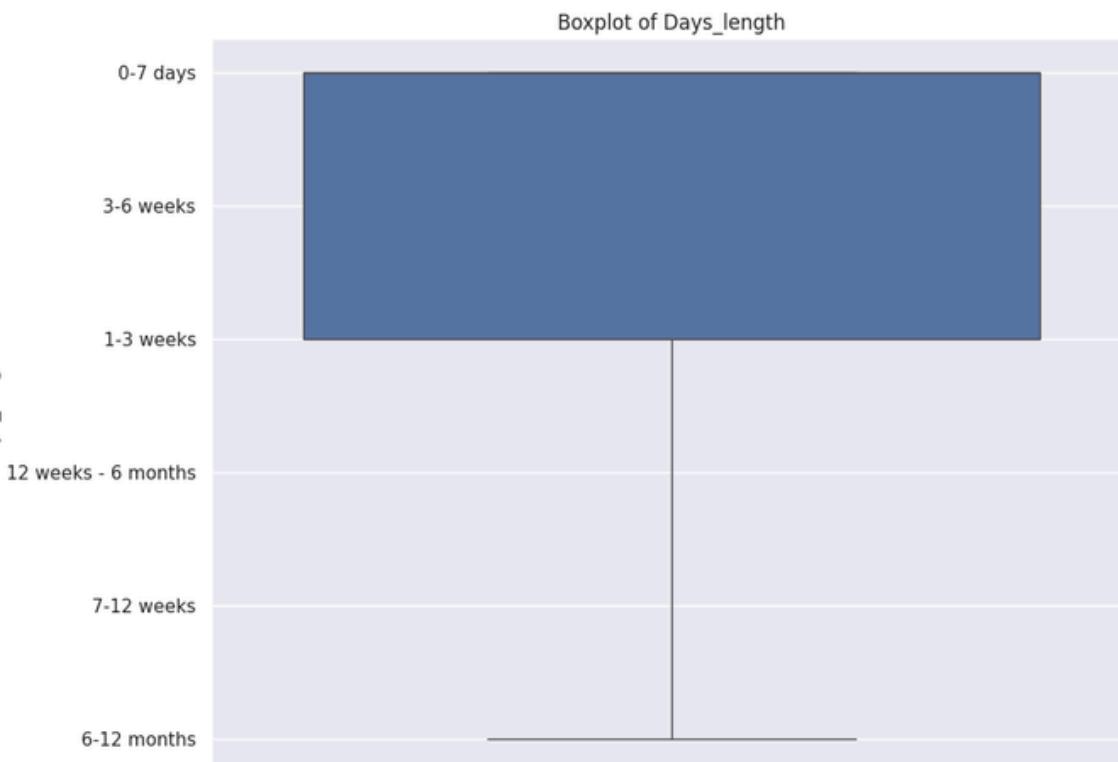
BOXPLOTS FOR NUMERICAL FEATURES

To check for outliers

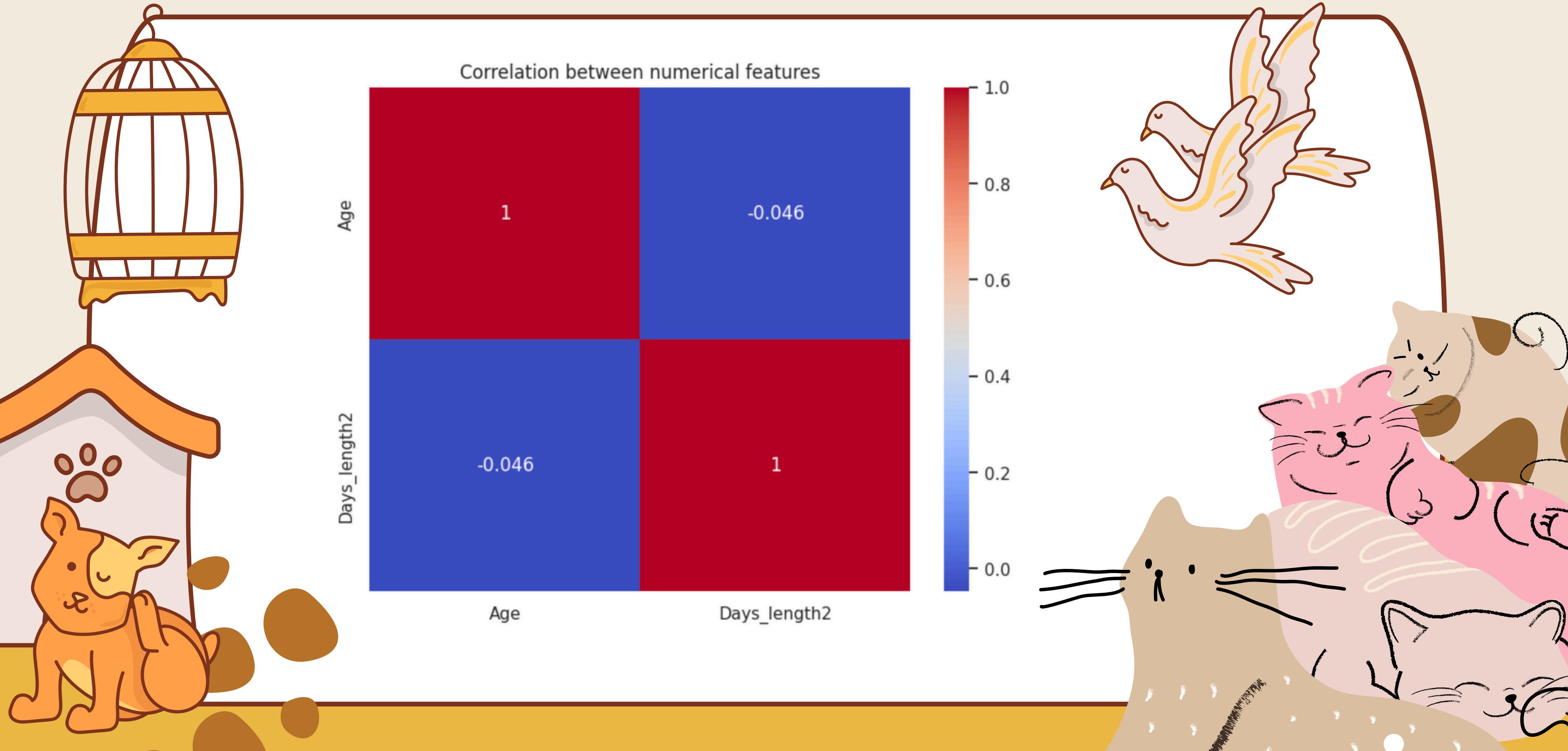
Age



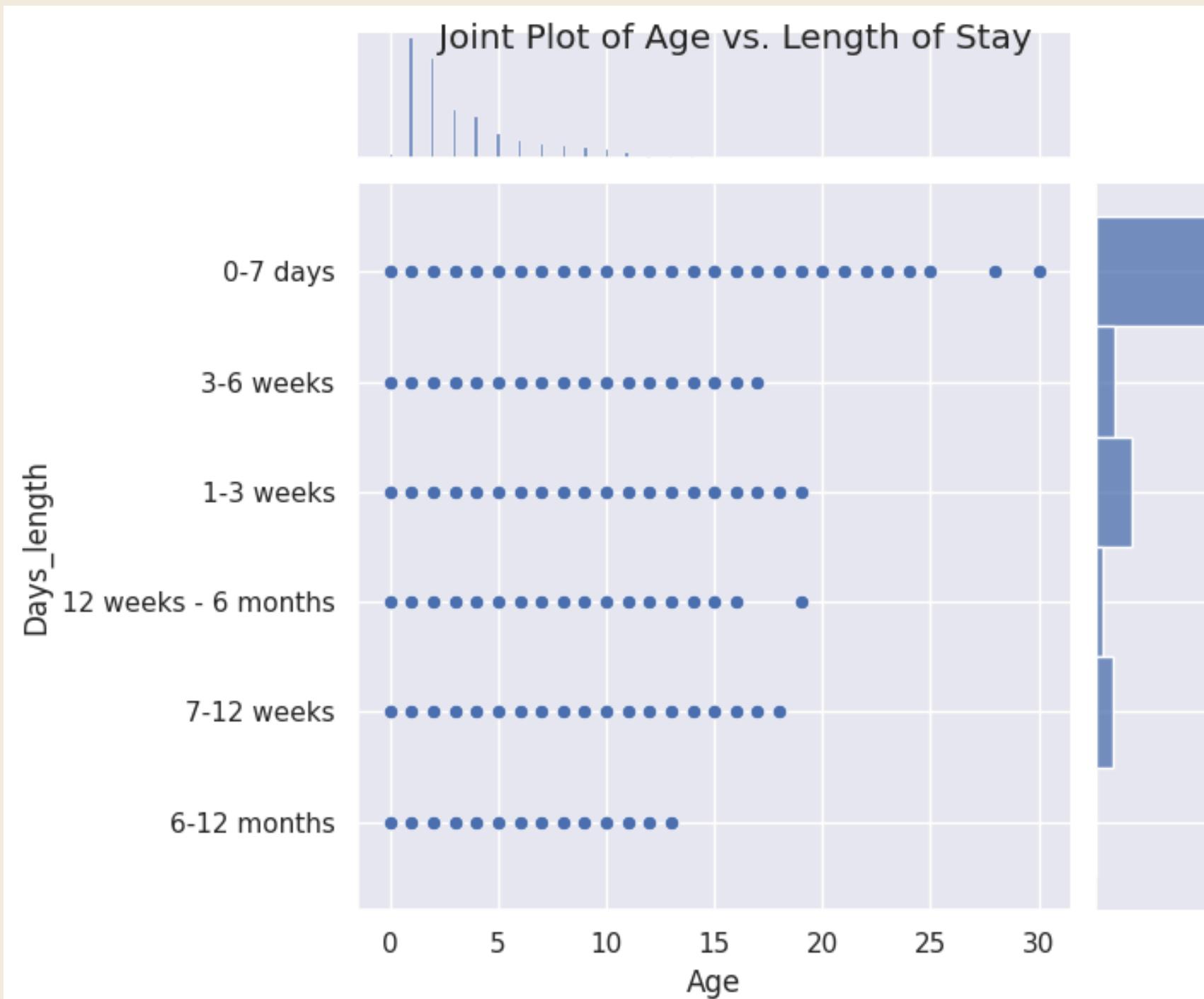
Day Length



CORRELATION OF NUMERIC COLUMNS



RELATIONSHIP BETWEEN AGE AND LENGTH OF STAY

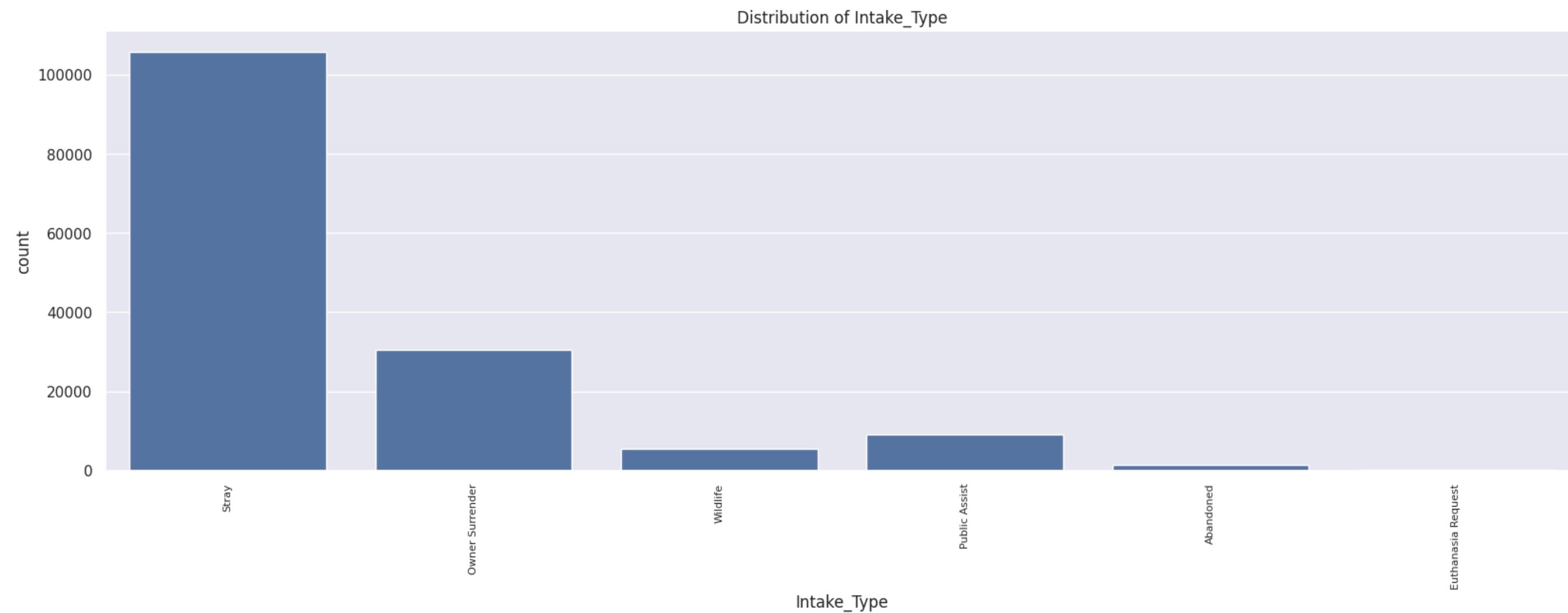


CATEGORICAL DATA



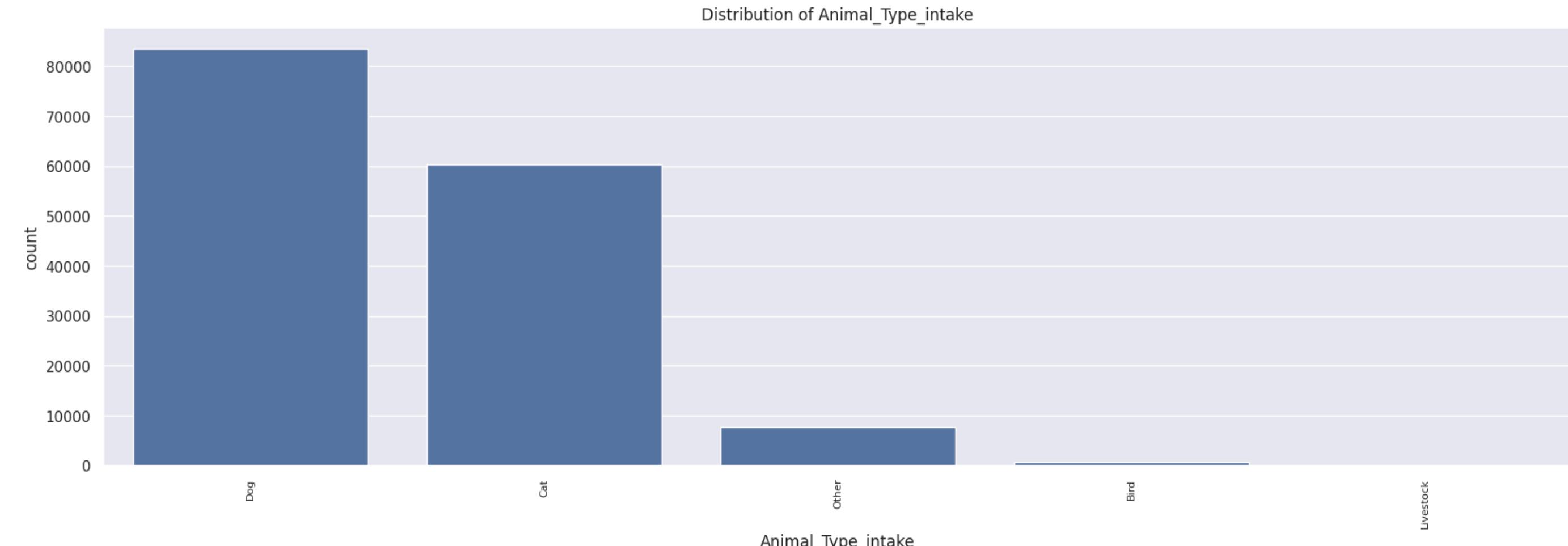
DISTRIBUTIONS OF CATEGORICAL FEATURES

Distribution of Intake Type



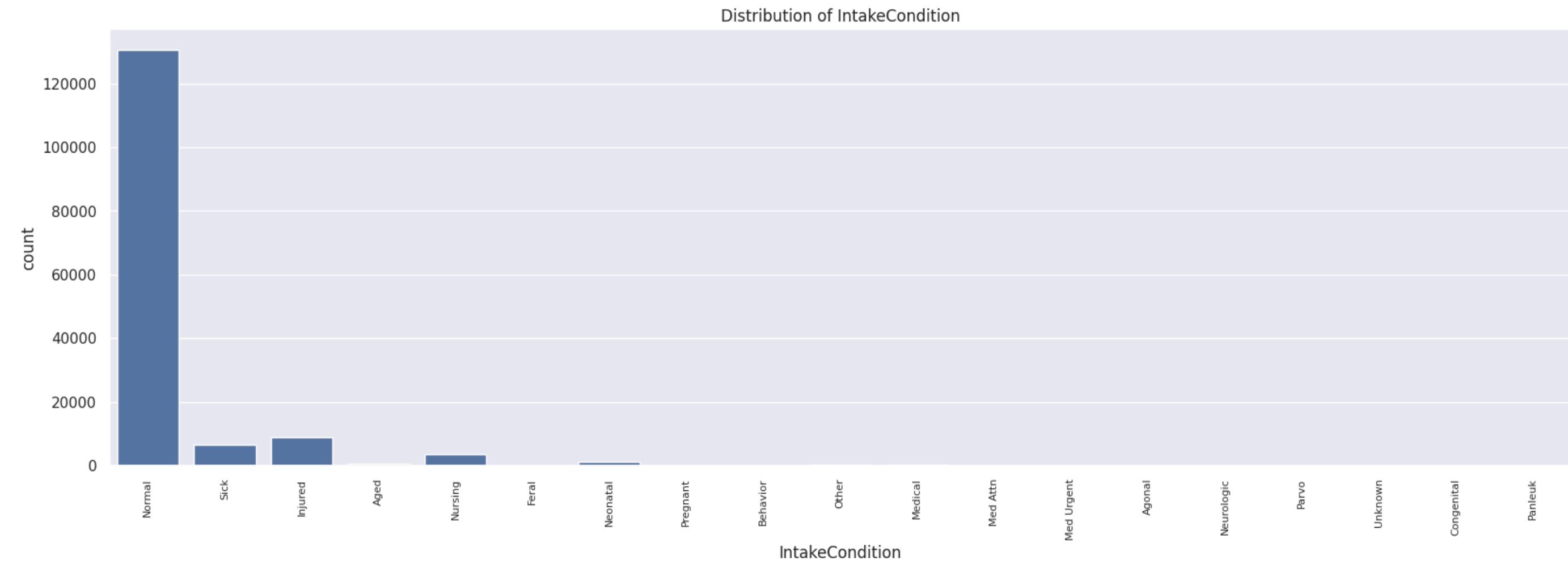
DISTRIBUTIONS OF CATEGORICAL FEATURES

Distribution of Intake Animal Type



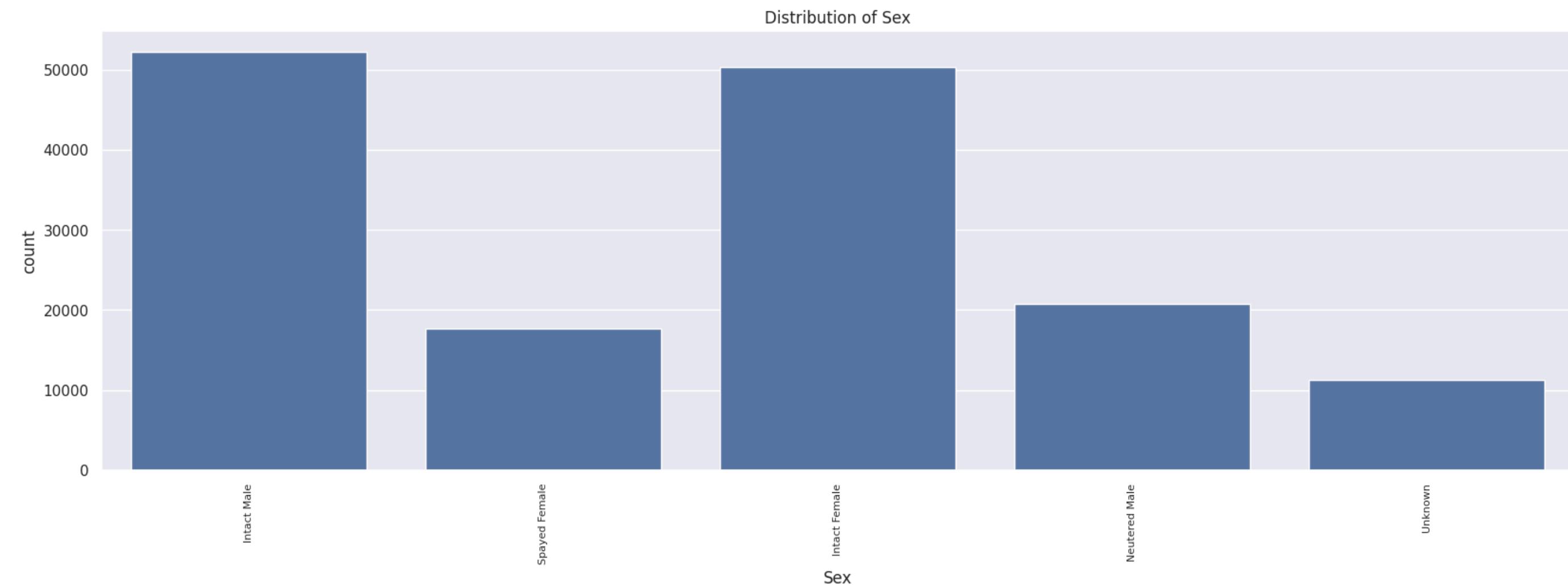
DISTRIBUTIONS OF CATEGORICAL FEATURES

Distribution of Intake Condition

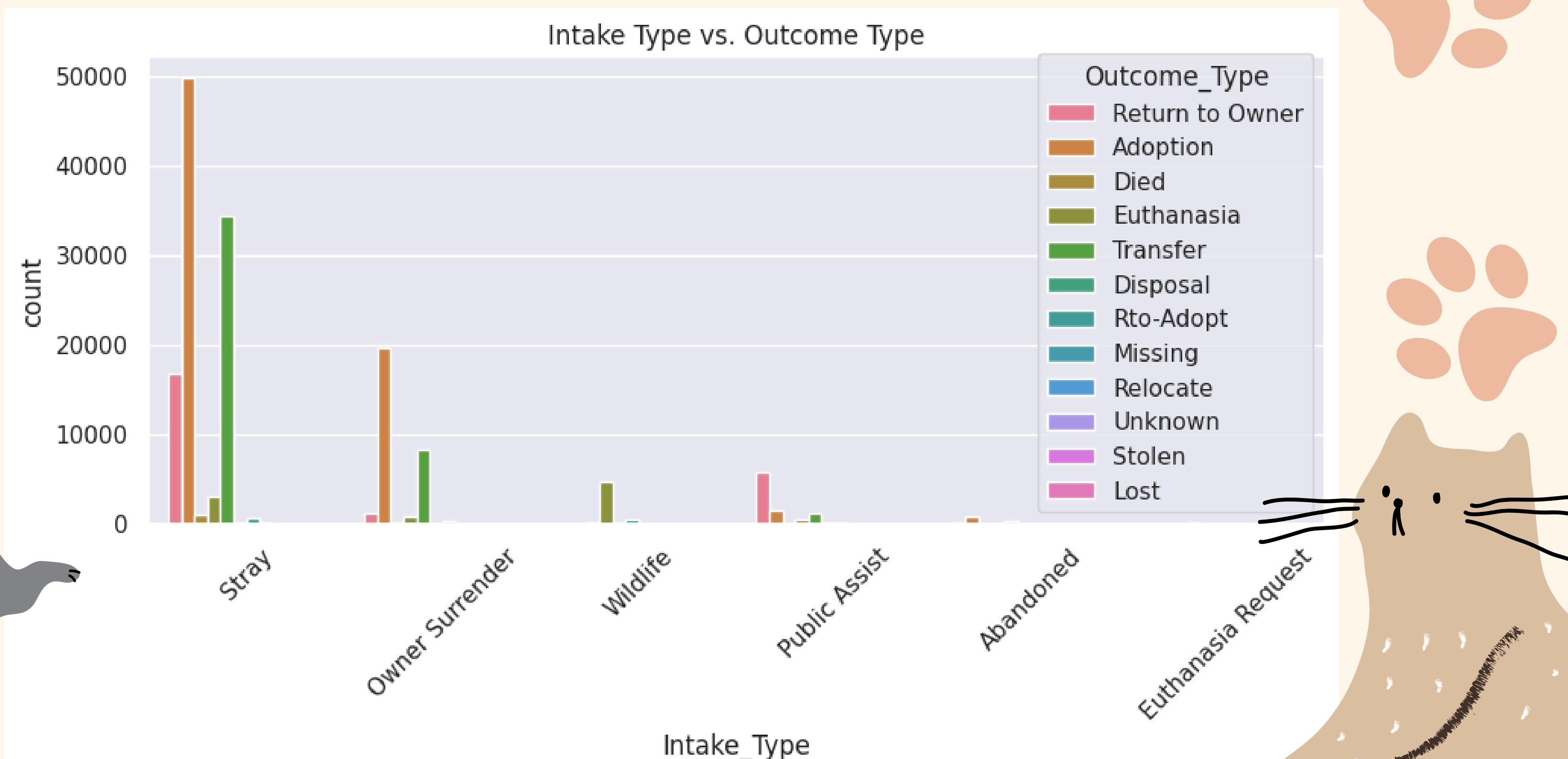


DISTRIBUTIONS OF CATEGORICAL FEATURES

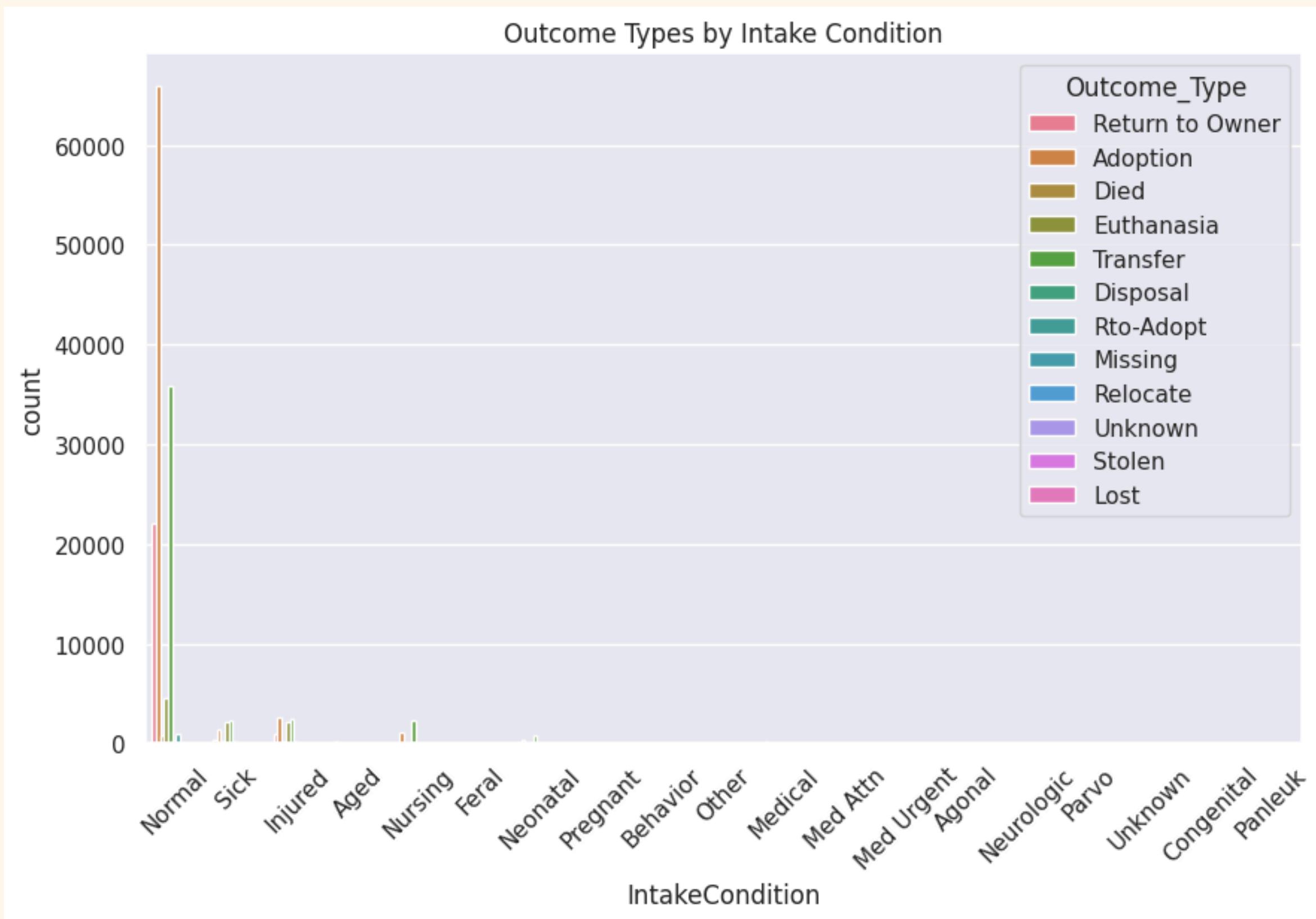
Distribution of Sex



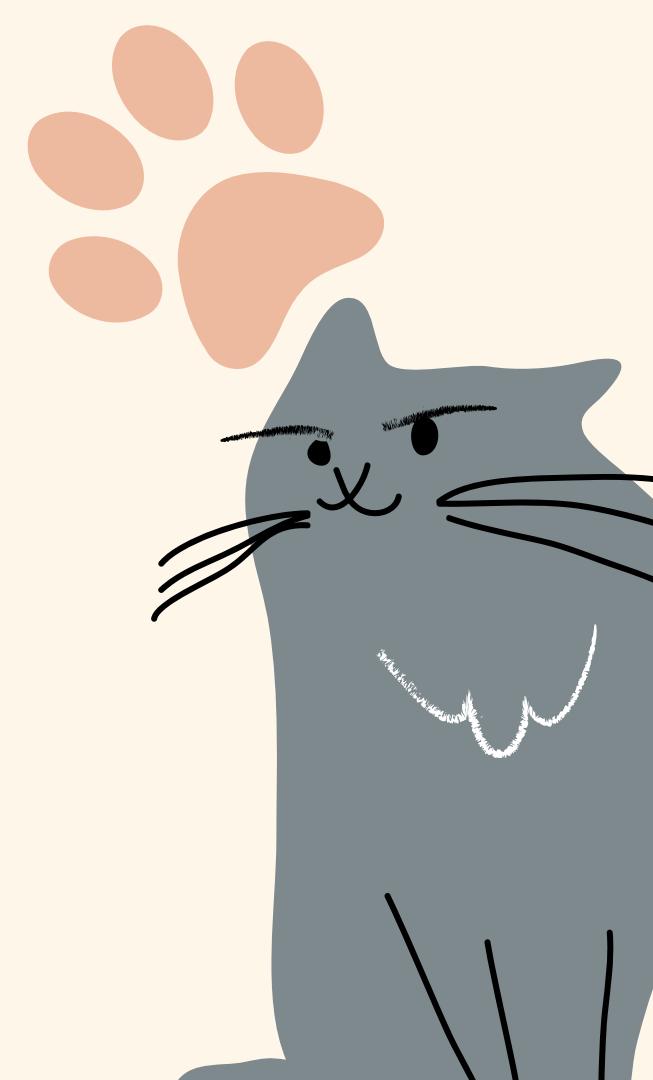
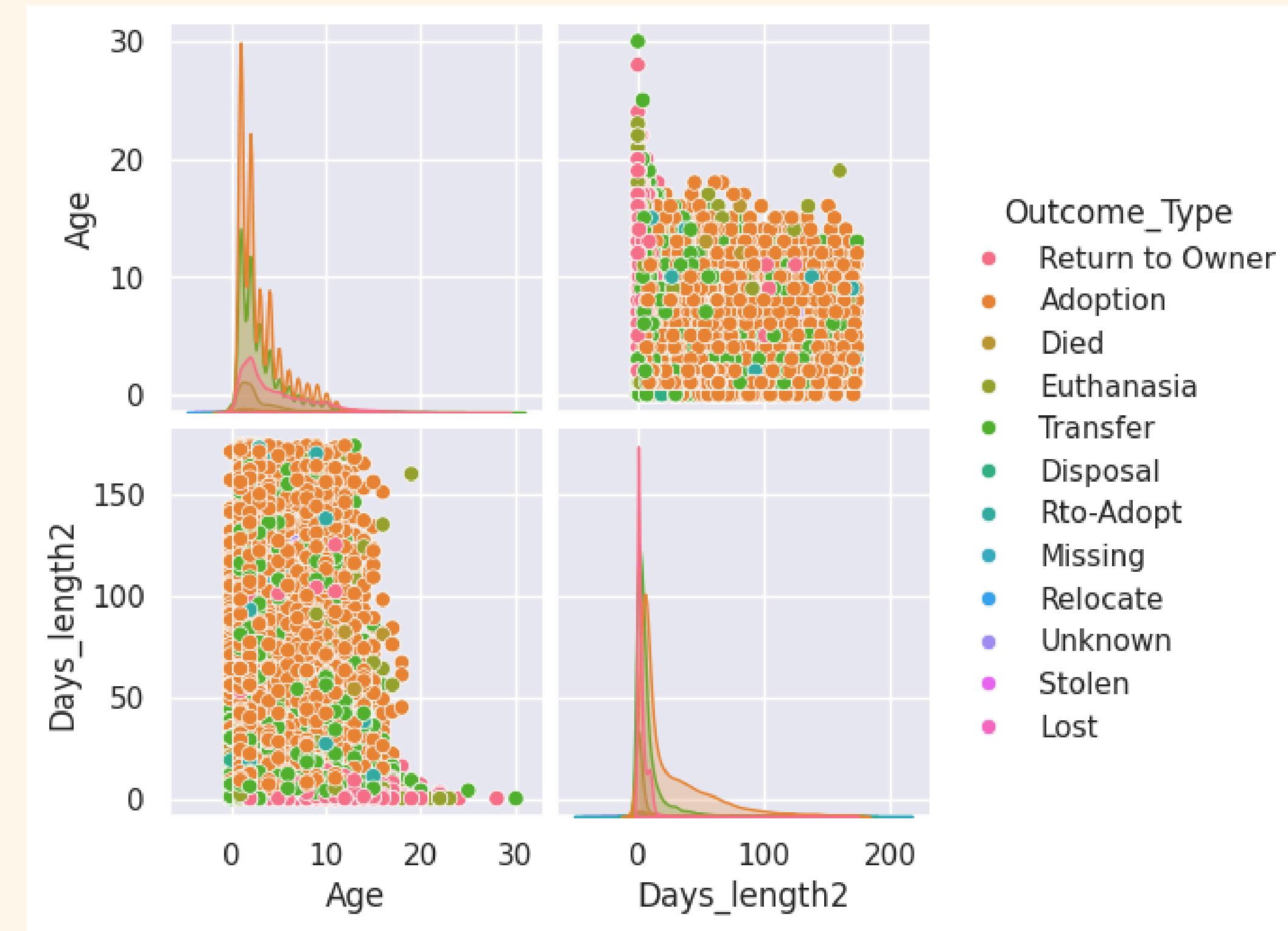
RELATIONSHIP BETWEEN INTAKE TYPE AND OUTCOME TYPE



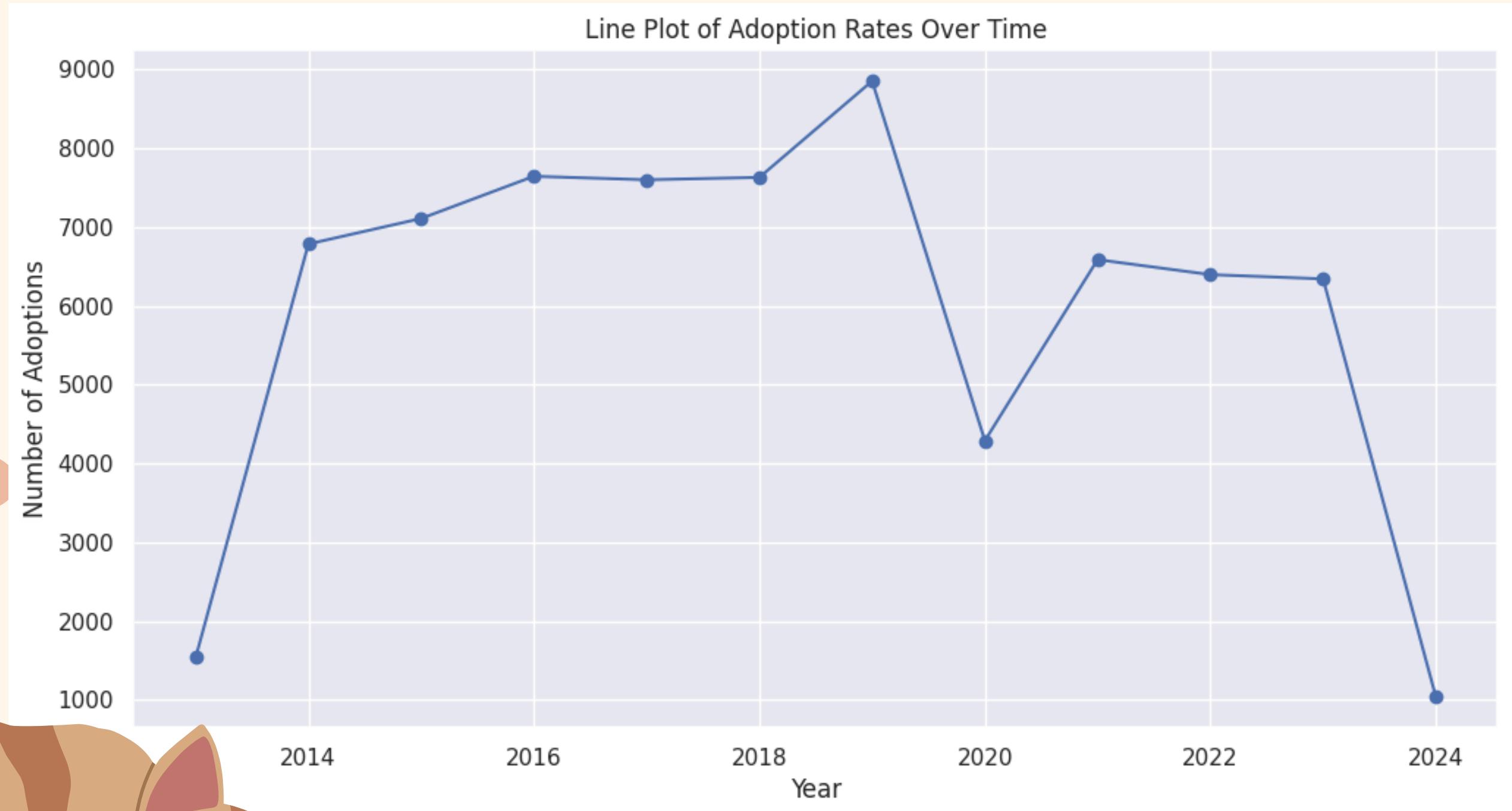
RELATIONSHIP BETWEEN INTAKE AND OUTCOME CONDITION



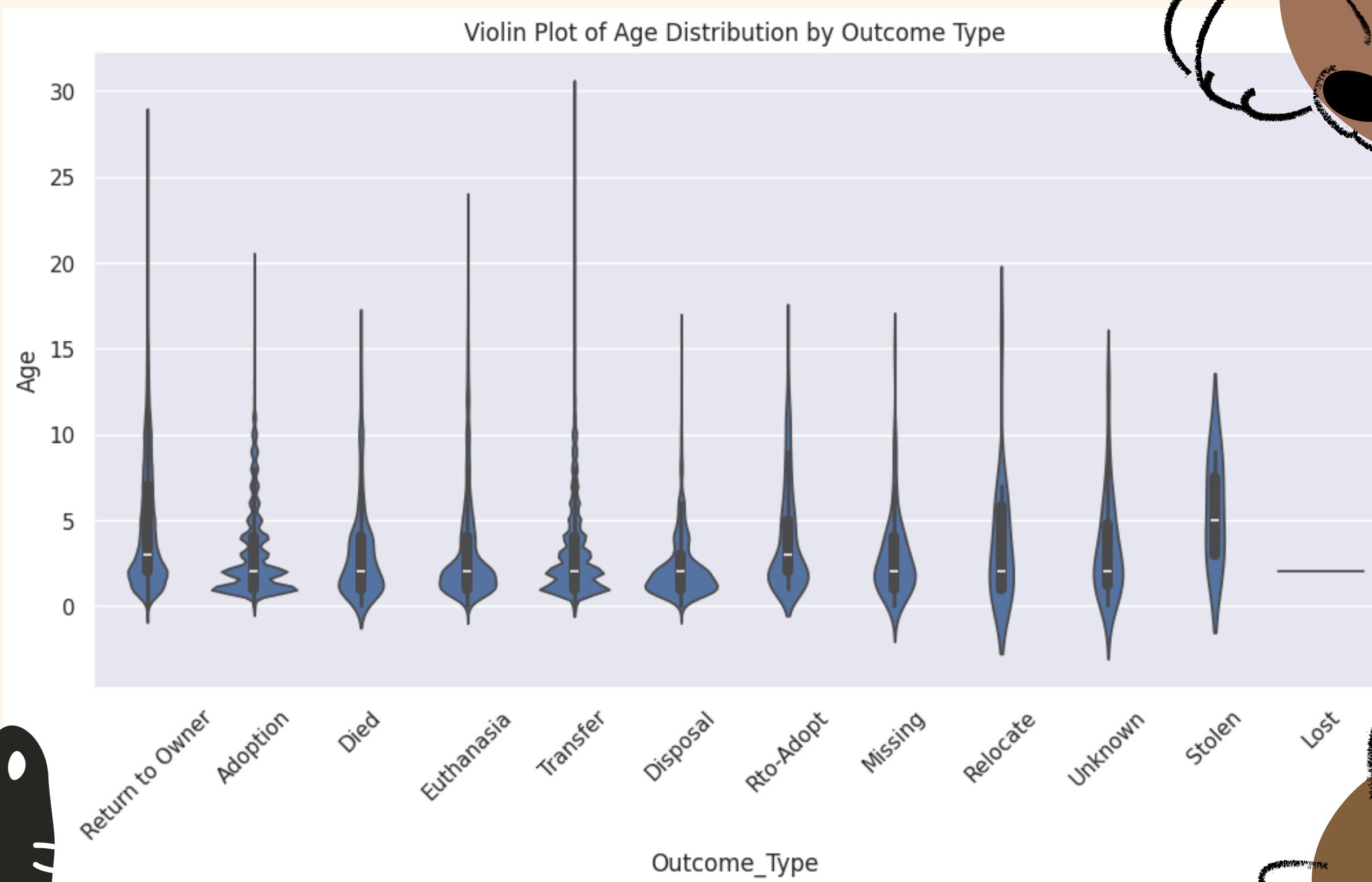
PAIRWISE RELATIONSHIPS IN THE DATASET



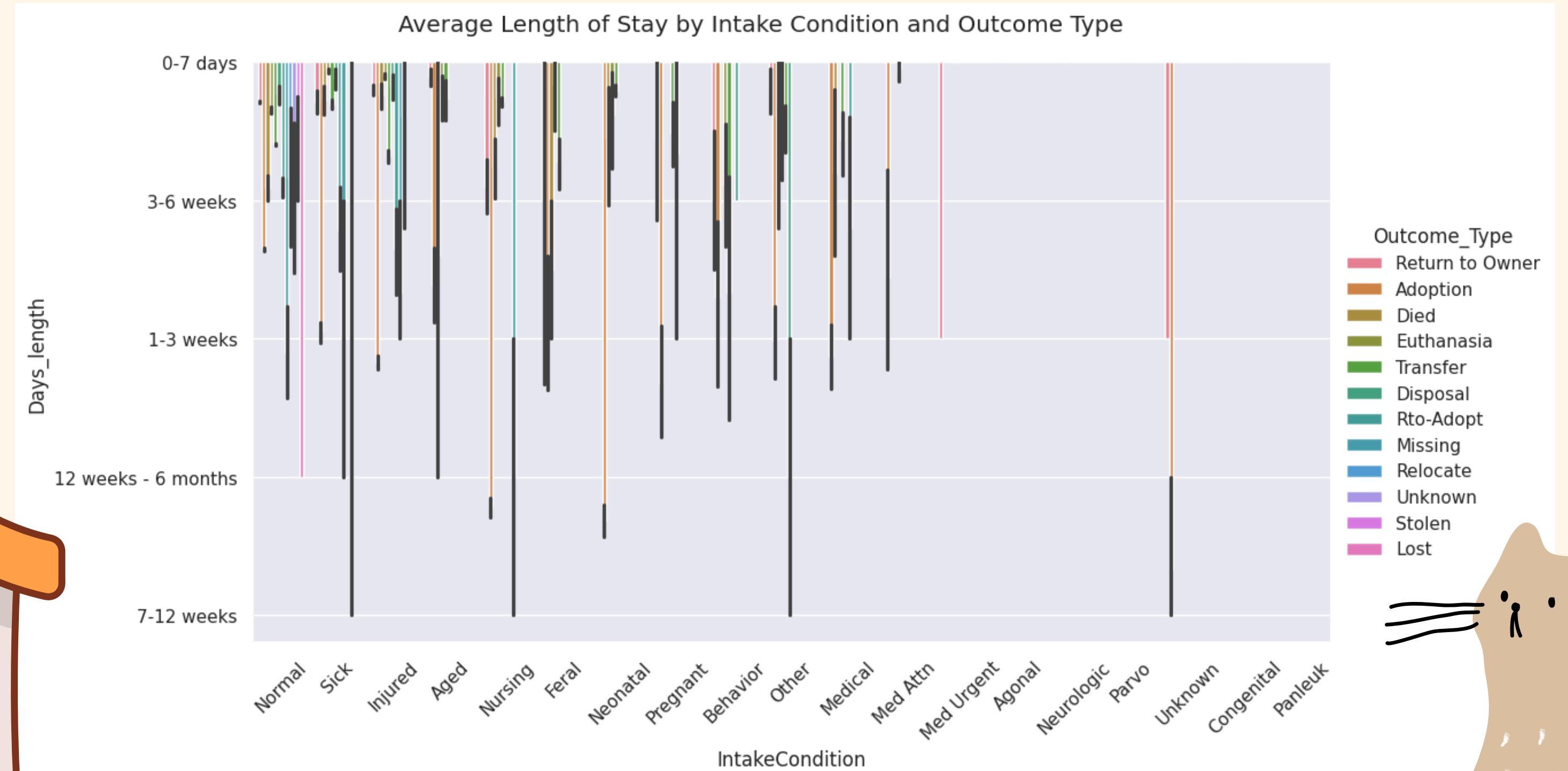
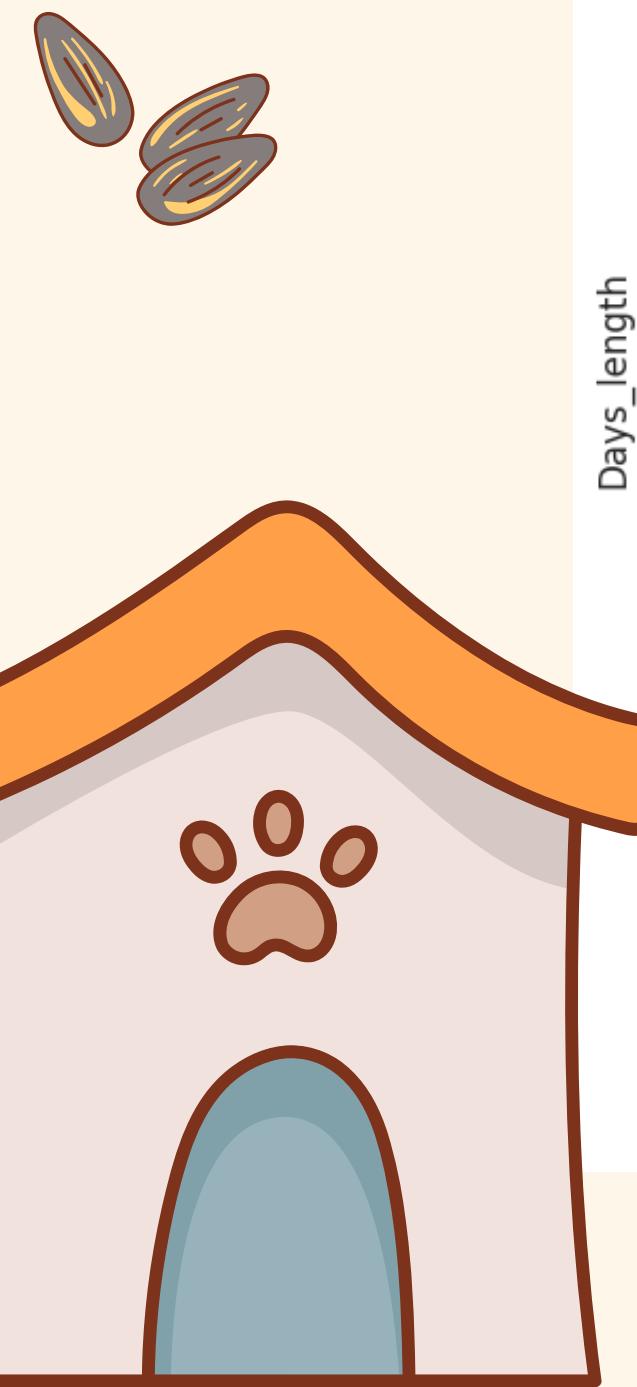
ANALYSING ADOPTION RATES OVER TIME



AGE DISTRIBUTION BY OUTCOME TYPE



LENGTH OF STAY AT THE SHELTER AND INTAKE CONDITION AND OUTCOME TYPE



MACHINE LEARNING

DEVELOPING ML MODELS:

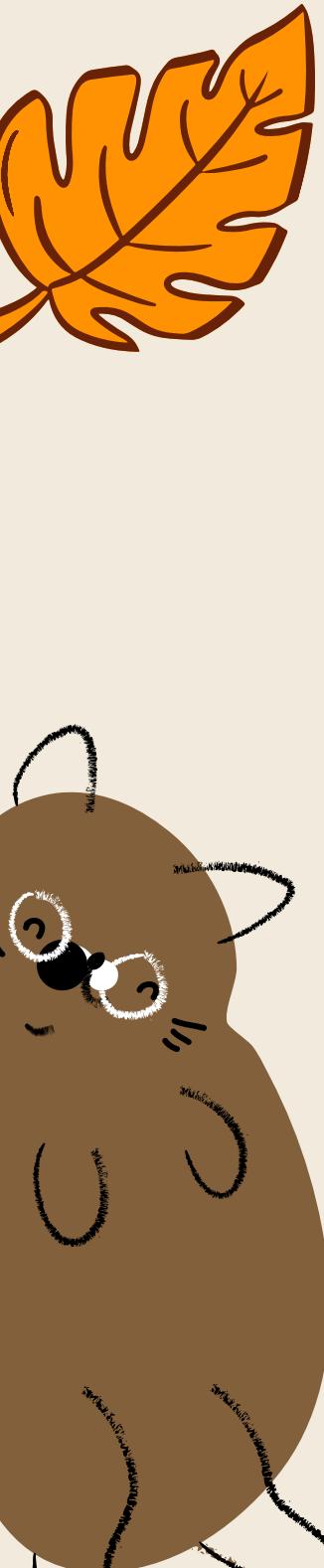
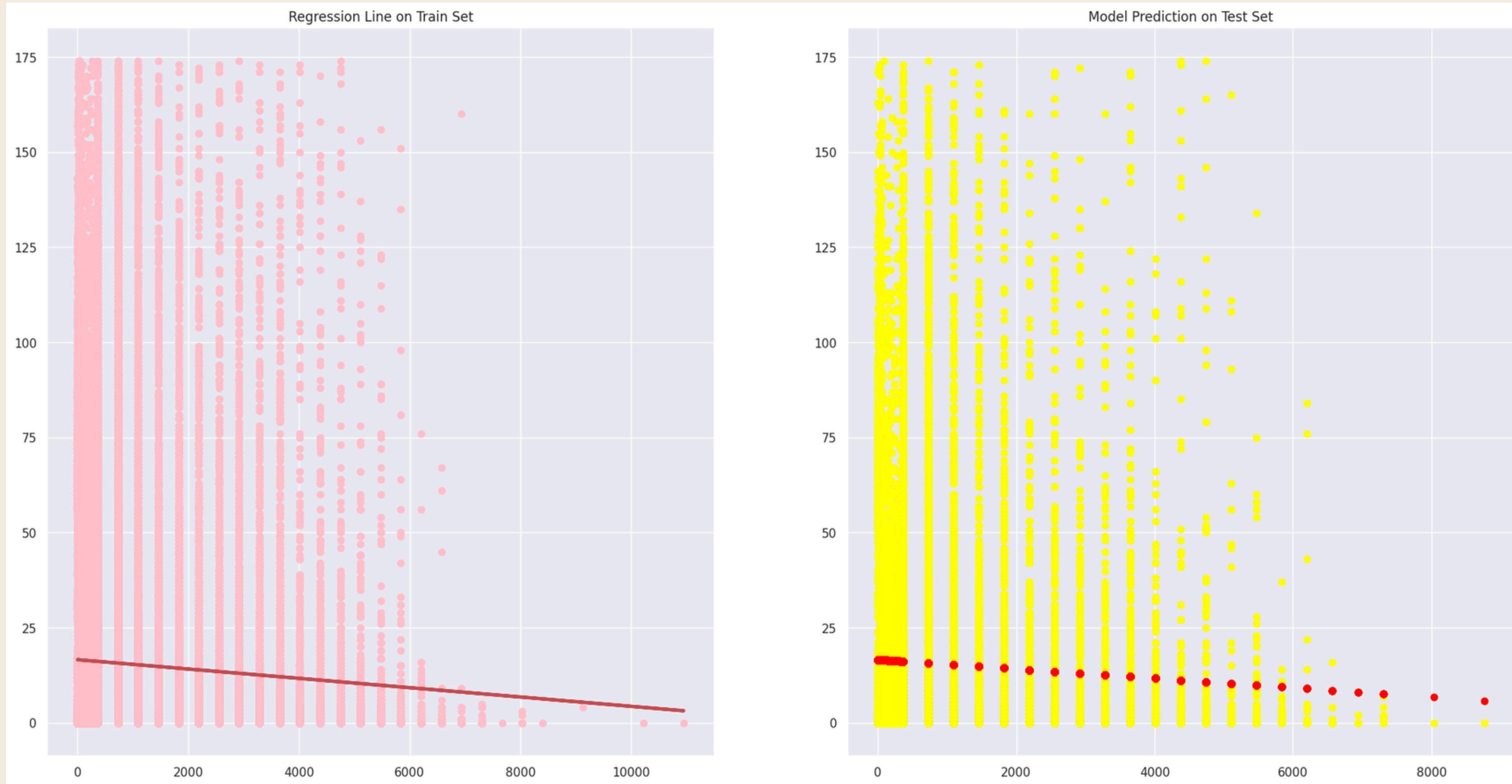
- USE NUMERICAL DATA:
 - WITH UNIVARIATE LINEAR REGRESSION
- USE NUMERICAL AND CATEGORICAL DATA:
 - GENERAL FEATURE SELECTION AND DATA PREPARATION
 - WITH DECISION TREES
 - WITH RANDOM FOREST
 - LASSO REGRESSION
 - XGBOOST



NUMERICAL DATA



1.1 UNIVARIATE LINEAR REGRESSION





USING NUMERICAL AND CATEGORICAL DATA





GENERAL FEATURE SELECTION & DATA PREPARATION

ANOVA Test



ANOVA test results:

ANOVA test results:

F-Value: 141.89141397989567 P-Value: 2.7362500761891012e-62

	group1	group2	meandiff	p-adj	lower	upper	reject
0	Aged	Agonal	-6.0412	1.0000	-69.7861	57.7038	False
1	Aged	Behavior	18.3952	0.0001	5.5964	31.1940	True
2	Aged	Congenital	-9.0412	1.0000	-99.0977	81.0154	False
3	Aged	Feral	4.8927	0.8997	-3.8346	13.6199	False
4	Aged	Injured	7.1471	0.0000	2.9562	11.3381	True
...
166	Parvo	Sick	10.6978	1.0000	-41.2548	62.6504	False
167	Parvo	Unknown	27.6000	0.9730	-29.2982	84.4982	False
168	Pregnant	Sick	-12.4044	0.0000	-20.1701	-4.6387	True
169	Pregnant	Unknown	4.4978	1.0000	-19.9694	28.9650	False
170	Sick	Unknown	16.9022	0.5136	-6.3529	40.1573	False

171 rows x 7 columns

Pairwise comparison results (Tukey's HSD):

Aged vs Agonal - Tukey HSD test statistic: -6.0412 P-Value: 1.0
Aged vs Behavior - Tukey HSD test statistic: 18.3952 P-Value: 0.0001
Aged vs Congenital - Tukey HSD test statistic: -9.0412 P-Value: 1.0
Aged vs Feral - Tukey HSD test statistic: 4.8927 P-Value: 0.8997
Aged vs Injured - Tukey HSD test statistic: 7.1471 P-Value: 0.0
Aged vs Med Attn - Tukey HSD test statistic: 0.9588 P-Value: 1.0
Aged vs Med Urgent - Tukey HSD test statistic: -7.8983 P-Value: 1.0
Aged vs Medical - Tukey HSD test statistic: 11.3516 P-Value: 0.0
Aged vs Neonatal - Tukey HSD test statistic: 10.5341 P-Value: 0.0
Aged vs Neurologic - Tukey HSD test statistic: -8.93 P-Value: 0.9999
Aged vs Normal - Tukey HSD test statistic: 5.6625 P-Value: 0.0002
Aged vs Nursing - Tukey HSD test statistic: 12.068 P-Value: 0.0
Aged vs Other - Tukey HSD test statistic: 8.6098 P-Value: 0.0005
Aged vs Panleuk - Tukey HSD test statistic: -9.0412 P-Value: 1.0
Aged vs Parvo - Tukey HSD test statistic: -10.0412 P-Value: 1.0
Aged vs Pregnant - Tukey HSD test statistic: 13.061 P-Value: 0.0
Aged vs Sick - Tukey HSD test statistic: 0.6567 P-Value: 1.0
Aged vs Unknown - Tukey HSD test statistic: 17.5588 P-Value: 0.4654
Agonal vs Behavior - Tukey HSD test statistic: 24.4364 P-Value: 0.9985
Agonal vs Congenital - Tukey HSD test statistic: -3.0 P-Value: 1.0
Agonal vs Feral - Tukey HSD test statistic: 10.9338 P-Value: 1.0
Agonal vs Injured - Tukey HSD test statistic: 13.1883 P-Value: 1.0
Agonal vs Med Attn - Tukey HSD test statistic: 7.0 P-Value: 1.0
Agonal vs Med Urgent - Tukey HSD test statistic: -1.8571 P-Value: 1.0
Agonal vs Medical - Tukey HSD test statistic: 17.3927 P-Value: 1.0
Agonal vs Neonatal - Tukey HSD test statistic: 16.5752 P-Value: 1.0
Agonal vs Neurologic - Tukey HSD test statistic: -2.8889 P-Value: 1.0
Agonal vs Normal - Tukey HSD test statistic: 11.7036 P-Value: 1.0
Agonal vs Nursing - Tukey HSD test statistic: 18.1092 P-Value: 1.0
Agonal vs Other - Tukey HSD test statistic: 14.6509 P-Value: 1.0
Agonal vs Panleuk - Tukey HSD test statistic: -3.0 P-Value: 1.0

One Hot Encoding

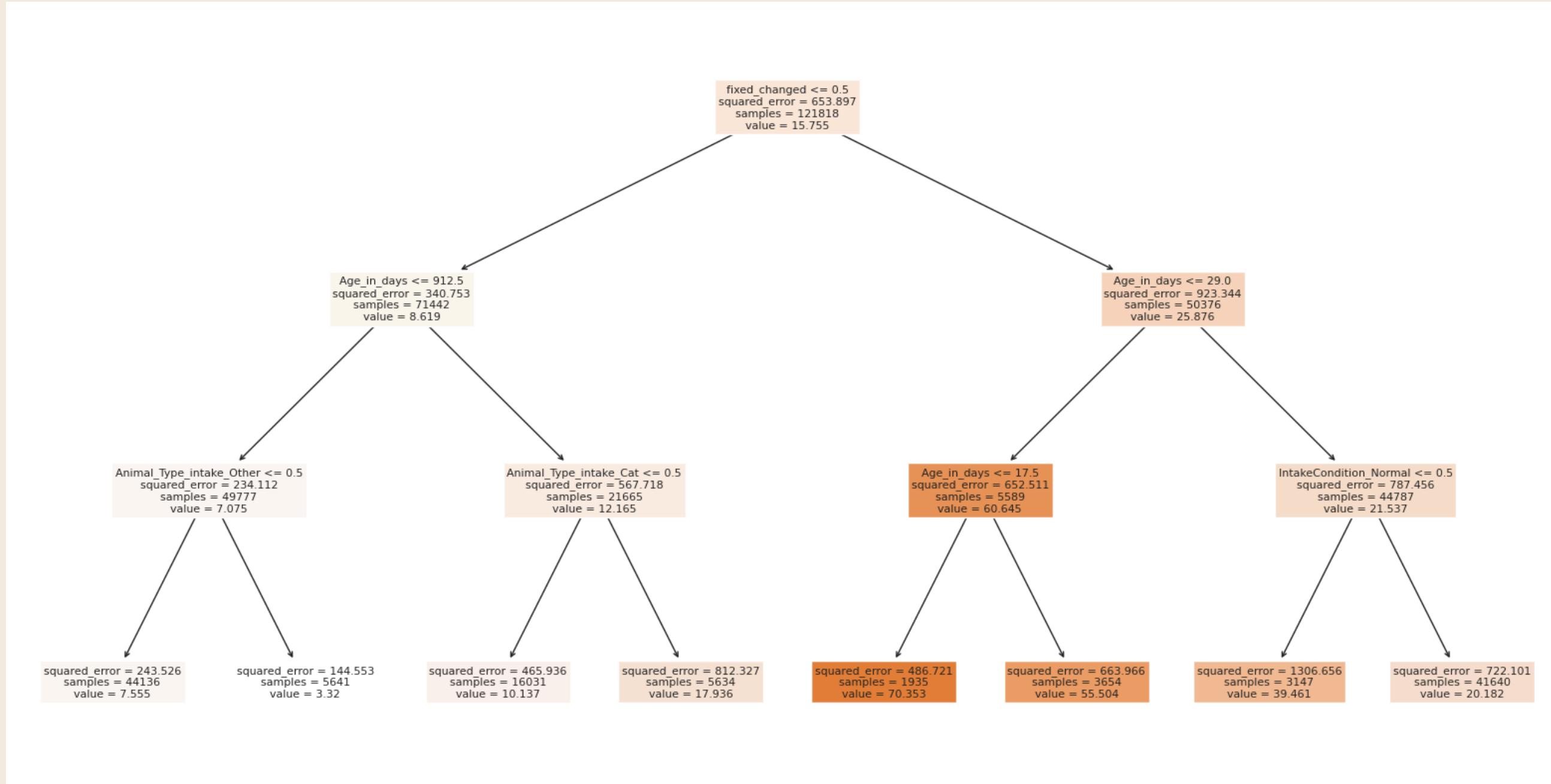
Outcome_Type_Return to Owner	Outcome_Type_Rto- Adopt	Outcome_Type_Stolen	Outcome_Type_Transfer
1.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0
1.0	0.0	0.0	0.0



DECISION TREE



Decision Tree Regression Model for Predicting Shelter Animal Length of Stay



	R^2 Score	MSE
Train Set	0.002408	652.475273
Test Set	0.001750	667.213387

Cross Validation

- *AVERAGE MSE ACROSS ALL FOLDS: 610.8284818801725*
- *AVERAGE R-SQUARED ACROSS ALL FOLDS: 0.049381482601313566*
- *STANDARD DEVIATION OF MSE ACROSS ALL FOLDS: 22.65856819779723*
- *STANDARD DEVIATION OF R-SQUARED ACROSS ALL FOLDS: 0.11586274740515004*

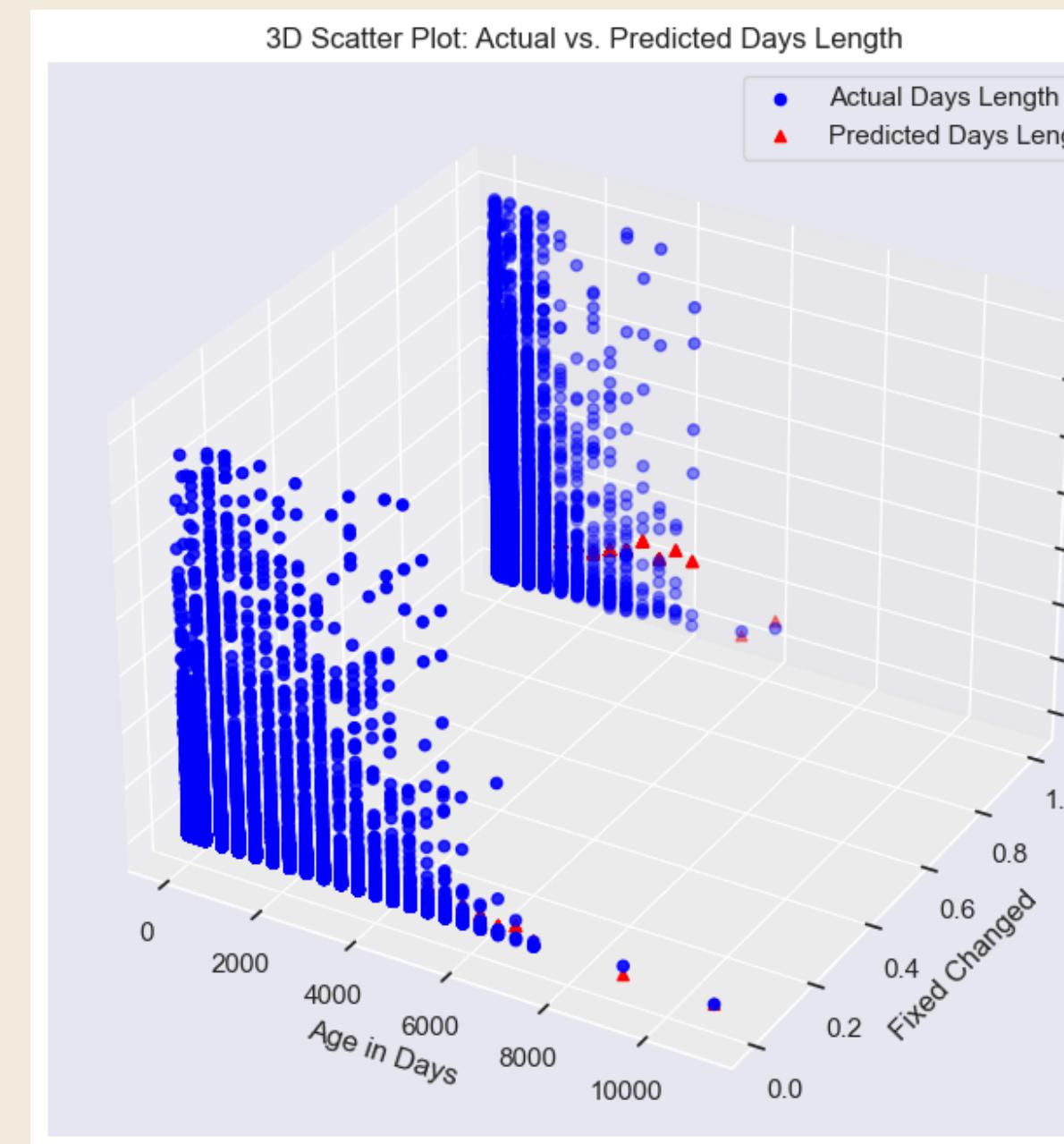
RANDOM FOREST



FEATURE IMPORTANCE ANALYSIS

	importance
Age_in_days	0.365390
fixed_changed	0.064246
Color_intake_Black/White	0.009812
IntakeCondition_Normal	0.008853
Color_intake_Brown	0.007184
...	...
Breed_intake_treeing walker coonhound/australia...	0.000000
Breed_intake_black mouth cur/australian cattle dog	0.000000
Breed_intake_pit bull/akita	0.000000
Breed_intake_scottish terrier/basset hound	0.000000
Color_intake_Yellow/Yellow	0.000000

SCATTER PLOT FOR ACTUAL VS. PREDICTED VALUES

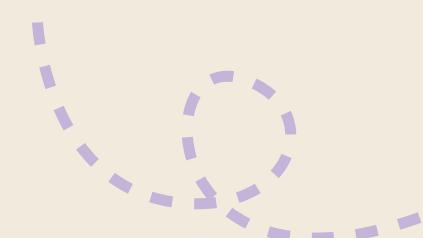


LASSO REGRESSION



Lasso Regression

- TRAINING MEAN SQUARED ERROR: 19.634127180597098
- TEST MEAN SQUARED ERROR: 19.569667740564995
- TRAINING R-SQUARED: 0.0
- TEST R-SQUARED: -5.319835968409237E-05
- COEFFICIENT
- AGE_IN_DAYS 0.0
- ANIMAL_TYPE_INTAKE -0.0
- BREED_INTAKE 0.0
-

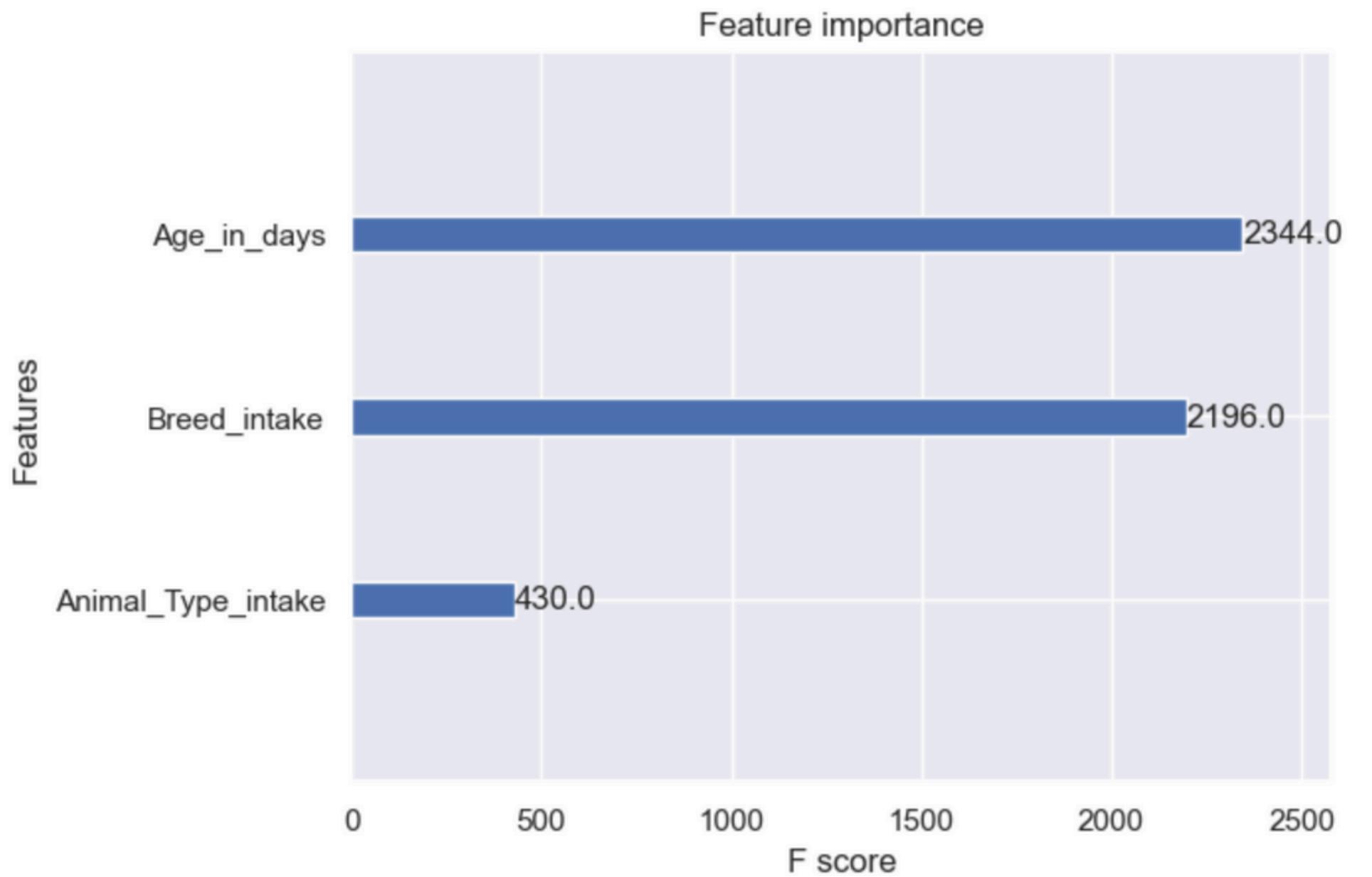




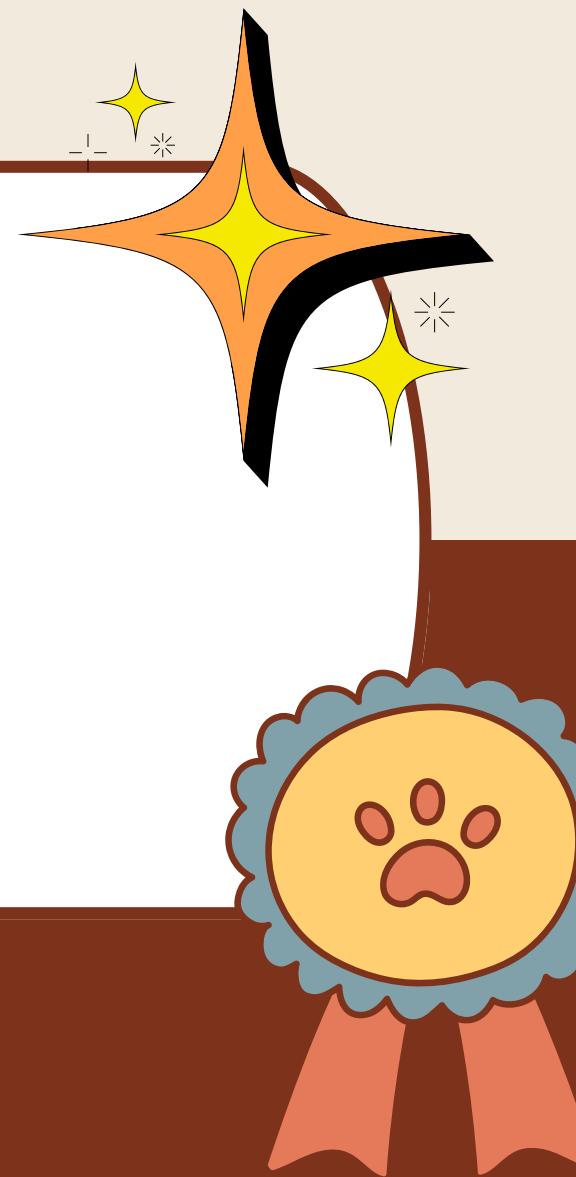
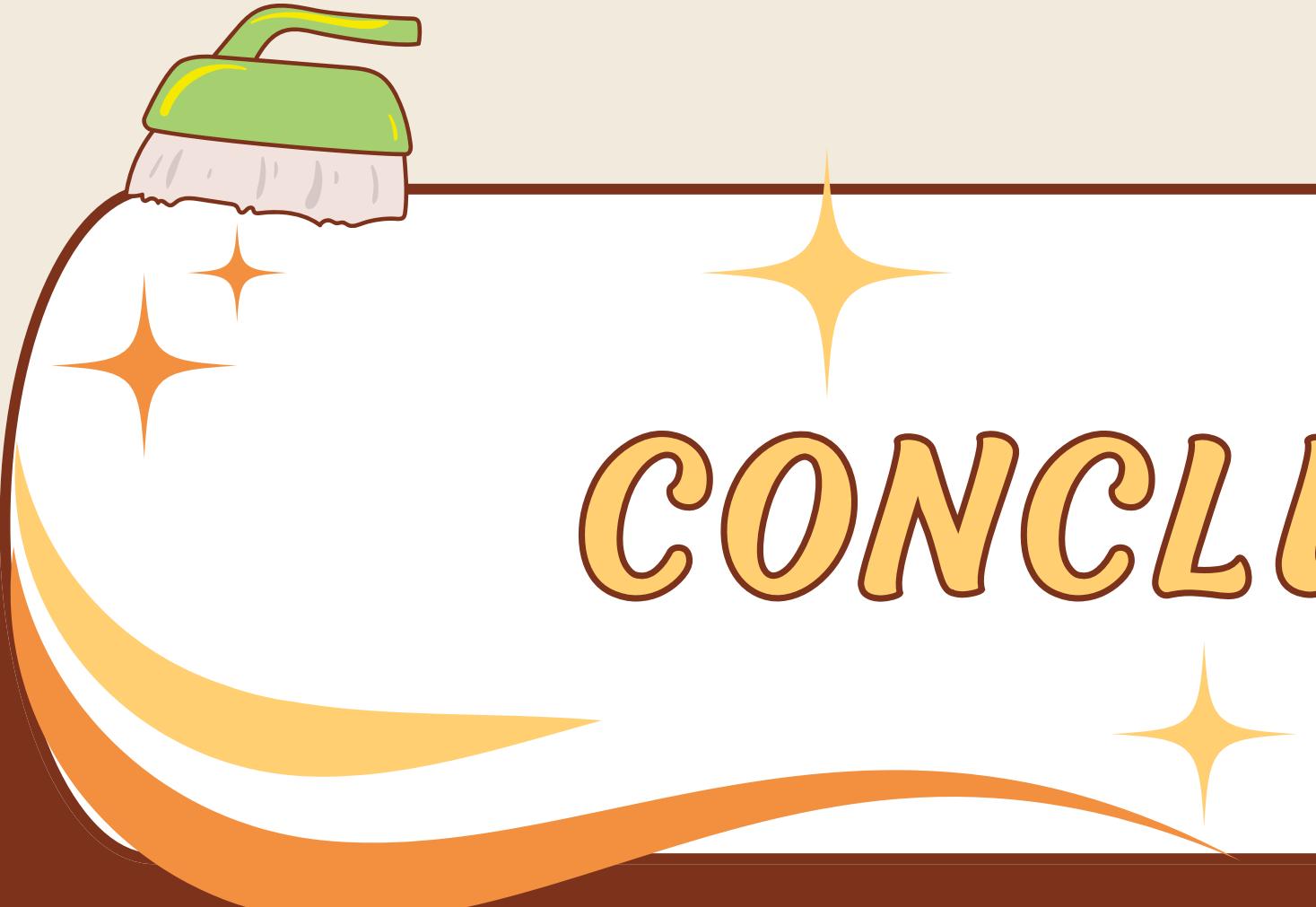
XGBOOST

XGBoost

MSE: 18.58
R-squared: 0.05



CONCLUSION



CONCLUSION

Major Findings:

- Age and intake condition are critical in determining adoption likelihood.
- Older and medically impaired pets face higher adoption challenges.

Strategic Recommendations:

- Implement proactive medical care and rehabilitation for at-risk animals.
- Initiate targeted adoption drives focusing on vulnerable pets.
- Engage the community through storytelling and advocacy.

Implications:

- Adopting these strategies can lead to better animal welfare and higher adoption success.



THANK YOU

SC1015 FCCA Team8

