

Handling Strategies for Data Irregularities.

Feature	Data Quality Issue	Handling Strategy
AnimalID	Not data preparation feature	Do nothing
Name_Intake	No useful information, missing values(~30%)	Remove column
DateTime_Intake	None	Do nothing
FoundLocation	Cardinality	Merge synonymous cardinalities
IntakeType	None	Do nothing
IntakeCondition	None	Do nothing
AnimalType_Intake	None	Do nothing
SexuponIntake	Missing Values (~9%)	Do nothing
AgeuponIntake	Negative Value (2 rows), Outlier	Imputation , do nothing
Breed_Intake	Cardinality	Merge synonymous cardinalities
Color_Intake	Cardinality	Merge synonymous cardinalities
DateTime_Outcome	None	Do nothing
DateofBirth	None	Do nothing
SexuponOutcome	Missing Values (~9%)	Do nothing
AgeuponOutcome	Outlier, 8 values smaller than intake age	Do nothing, drop rows
binary_outcome	None	Do nothing

Cleaned data results:

Continuous Features

	count	mean	std	min	25%	50%	75%	max
AgeuponIntake	992.0	780.548387	1105.729441	0.0	62.00	365.0	783.25	7670.0
AgeuponOutcome	992.0	800.413306	1104.058498	0.0	87.75	368.0	910.25	7671.0

Categorical Features

	count	unique	top	freq
DateTime_Intake	992	12	6	103
FoundLocation	992	14	Austin	839
IntakeType	992	5	Stray	719
IntakeCondition	992	8	Normal	870
AnimalType_Intake	992	5	Dog	534
SexuponIntake	992	5	Intact Male	337
Breed_Intake	992	12	Others	295
Color_Intake	992	9	Bi color	407
DateTime_Outcome	992	12	8	102
DateofBirth	992	12	4	128
SexuponOutcome	992	5	Neutered Male	337
binary_outcome	992	2	0	905