(1). Prepare a Data Quality Report Initial Findings

The dataset we work with is a sample of the data released by an animal shelter. The aim is to predict the binary outcome of an animal in the shelter, for this variable, the value “1” indicates that after intake, the animal outcome was negative, e.g., "death", while the value "0" indicates that the animal outcome was a positive one, e.g., was adopted or returned to the owner.

Based on examining the data in a spreadsheet program, 6 of the features are continuous and the rest categorical. 3 of these continuous values are datetime values which will be considered differently than the rest

- *Binary outcome* will be converted to **categorical** because it contains only 2 possible values, 0 or 1.
- *Age upon Intake* column will be converted to **continuous** and the values will represent the value in days to make the data uniform and comparable
- *Age upon Outcome* column will be converted to **continuous** and the values will represent the value in days to make the data uniform and comparable
- *‘MonthYear_Intake’, MonthYear_Outcome, DateTime_Intake, DateTime_Intake, Date of birth* are all converted to datetime data types
- All other object values will be converted to **categorical** values

Categorical Data - Reviewing the categorical data below we can see all unique values > 1

- *Animal ID* has all unique values (1000)
- *Found location* has 784 unique values
- *Intake Type* has 5 unique values
- *Intake Condition* has 7 unique values
- *Sex Upon Intake* has 5 unique values
- *Breed Intake* has 211 unique values
- *Color_Intake* has 104 unique values
- *Name_Outcome* has 638 unique values
- *Animal Type_Outcome* has 4 unique values
- *Sex Upon Outcome* has 5 unique values

Continuous Data - Reviewing the continuous data below we can see all have a non-zero standard deviation.

- This implies that a particular feature does not contain a single constant value in all of the rows. Thus in this case, none of the continuous features are constant.
- Result - No constant columns