Datasets Information

* = included columns

AAC\_Found.csv

* Source: <https://data.austintexas.gov/browse?q=dogs&sortBy=relevance>
* Columns:
  + Animal ID\*
  + Found Location
  + At AAC\*
  + Intake Date\*
  + Type (filter for dogs)
  + Looks Like\*
  + Color\*
  + Sex\*
  + Age\*
  + Image Link\*

AAC\_Intake.csv

* Source: <https://data.austintexas.gov/browse?q=dogs&sortBy=relevance>
* Columns:
  + Animal ID\*
  + Name
  + DateTime
  + MonthYear
  + Found Location\*
  + Intake Type\*
  + Intake Condition\*
  + Animal Type (filter for dogs)
  + Sex upon Intake\*
  + Age upon Intake\*
  + Breed\*
  + Color\*

AAC\_Outcome.csv

* Source: <https://data.austintexas.gov/browse?q=dogs&sortBy=relevance>
* Columns:
  + Animal ID\*
  + Name\*
  + DateTime\*
  + MonthYear
  + Date of Birth\*
  + Outcome Type\*
  + Outcome Subtype
  + Animal Type (filter for dogs)
  + Sex upon Outcome\*
  + Age upon Outcome\*
  + Breed
  + Color

Catsvdogs.csv

* Source: <https://data.world/datanerd/cat-vs-dog-popularity-in-u-s/workspace/file?filename=catsvdogs.xlsx>
* Columns:
  + Location\*
  + Number of Households (in 1000)\*
  + Percentage of households with pets\*
  + Number of Pet Households (in 1000)\*
  + Percentage of Dog Owners\*
  + Dog Owning\*
  + Mean Number of Dogs per household\*
  + Dog Population (in 1000)\*
  + Percentage of Cat Owners
  + Cat Owning
  + Mean Number of Cats per household
  + Cat Population

Data Cleaning Notes;

* Imported CSVs (AAC\_Found, AAC\_Intake, AAC\_Outcome, catsvdogs)
* Removed rows that referenced animals that were not dogs
* Created data frames from chosen columns for each
* Split sex column into gender(male/female) and sex(spayed, neutered, etc.) on found and intake data frames
* Split Location into street, city and state in the intakes data frame, then zip code, latitude and longitude in the found data frame
* Completed location data in found data frame by merging w/ intake data frame and removing unnecessary columns
* Created date\_id column in intakes data frame and outcomes data frame to differentiate multiple encounters with the same dog
* Merged intake and outcome data frames by animal ID and date\_id
  + Split color into 2 columns (color 2, color2) to separate entries with more than 1 color
  + Split breed into 2 columns (breed1, breed2) to separate entries with more than 1 breed or entries including “Mix”
  + Calculated a duration of stay column by subtracting intake date from outcome date
  + Removed rows with negative duration
  + Replaced nan values with “none” for color2, breed2 and name
  + Removed remaining rows with nan values
* renamed and reordered columns
* Exported 3 final data frames to CSVs