Data Cleaning Challenges

Week 1

* ~~Color column-split, sort and combine~~
* ~~Merge Intake and Outcome CSVs- Michelle~~
* ~~Breed column- split and categorized-Michelle working on~~
* ~~Add Mix column for Breed- Anna~~
* ~~Age- Bin into groups- Jordan~~
* ~~Sex- Column for Male/Female, Column for Spayed/Neutered- RoseAnne~~
* ~~Add length of stay column after merge~~
* Add size column based on Breed from AKC groups- coming from second database
* ~~Create column for multiple intakes (is this a yes no? i.e. multiple intakes 1 for yes and 0 for no) Are we still doing this?? Is this necessary- Discuss Wednesday~~

Week 2

Monday 1/24

* Data ETL still to do (Michelle)
  + ~~Color column~~
  + ~~Breed columns- second file reference (multiple columns)~~
  + ~~Length of stay~~
  + ~~Size -second file reference~~
* Database (RoseAnne)
  + ~~Create tables in SQL (3 tables)~~
* ML Model (Anna)
  + ~~Dummy coding~~
* Presentation/Visualizations
  + St~~art template in Google Slides~~
  + Visualizations from merged CSV
  + ~~Map out what presentation looks like- Flask vs. Tableau~~
* Wed Class- know required deliverables for segment 2
* Sat meeting- try to finalize tasks for the week, review to make sure project is up to date with segment 2, plan for week 3.

Saturday 1/29

* ~~Data ETL- color column, breed restriction column~~
* ~~Database- create tables (5 tables- intakes, outcomes, breed list, breed restrictions, final merge)~~
* ~~ML Model- finalize dummy coding, start drafting model~~
* ~~Ask about DB connection requirements for project~~
* Target dates: 2/6- finalize DB and model, 2/9- review any changes, 2/12- finalize dashboard. Final Presentation 2/16

Wednesday 2/3

* Add CSV files to main branch
* Add data cleaning code to main branch
* Add ML code to main branch
* Add descriptions of Cleaning, Database set up, and ML model to README
* Review and add to google slides
* Storyboard for Dashboard

2nd Segment To do’s

* Database integration
* ML model
* Start Dashboard presentation (Module says Google Slides)

Additional questions of the data, or future evolution of the project

* Can we determine which factors have the most impact on adoptability
* Can we predict length of stay in the shelter using the features from the dataset?
* Does fostering have any impact on adoptability? – might be better for after analysis is complete if time permits.
* How does covid impact animal adoptions- split the data pre-pandemic vs pandemic adoptions

Goals of the Analysis

* Increasing adoptability
* Decreasing euthanizing when possible (not due to health or behavioral concerns) \*\* dataset is from no kill shelter, so the data might not help with this.
* Improve use of shelter resources

Additional info to find for Presentation

* # of animals in shelters, percentage adopted and euthanized- nationwide

Presentation Visualizations

* Tableau
* Pandas- pyplot
* Webpage dashboard- time permitting