Project 2 - Group 14

*Group members*: **Jana, Tia, Anthony**

*Git repo*: [**git@github.com:AGGII-52/Project-2\_Group-14.git**](mailto:git@github.com:AGGII-52/Project-2_Group-14.git)

*Datasets*:

1.Rowan County Adoptable Animal List (Dogs, Cats) **https://docs.google.com/spreadsheets/d/1Lhjn0vXGAFhgntoZJ7BuUNBBsd\_zU81YQYLXA7G96iA/edit?fbclid=IwAR3P1pobzc1ml3bcYiHS1dUohkpaBI-rsTse3erJfe0U88Fmi09SG5w1QNs#gid=2**

2.SoCo Data

**https://data.sonomacounty.ca.gov/Government/Animal-Shelter-Intake-and-Outcome/924a-vesw**

We used the above sources based on these data sets availability and their ability to list animals currently available for adoption.

● Detailing the process of the extraction, transformation, and loading steps

Extract:

Tia and Jana pulled the CSV files from the sources. We dropped the Petfinder API due to time constraints and having two sets of data from different sources. Tia read the three CSV files into a Jupyter notebook “Cat and Dog Extract File.ipynb”.

● Explain why you have performed the types of transformation you did

Transform:

After viewing the data, Tia found the columns of data we wanted to view, “Animal ID”, “Description/Breed”, “Sex”, and “Intake”. The Rowan County data listed animals by description and the SoCo data listed breed, we determined it would be best for our final database to rename these columns to a combined “Description/Breed” column.

We then determined both data sets list whether an animal was a stray, surrendered, or seized. Again, to unify our final database, Tia renamed the SoCo data “Intake Subtype” to match the Rowan County data label of “Intake”.

Jana took the cleaned df for each CSV, after combining the data frames she added a column to identify the animal’s shelter. The combined data was loaded into a Postgres SQL database.

  

● Why you chose the type of final database

Relational database, because there are two different types of data and both are data sets pertaining to animals.

● Schema of the tables/collections in the final database

Load:

Anthony

● Hypothetical use case(s) for your database

*Use*:

To determine

1.Which species of animal is in most need of adoption/has the largest numbers available?

2.Which breed of animal is in most need of adoption/has the largest numbers available?