**Project Report**

* **E**xtract: your original data sources and how the data was formatted (CSV, JSON, pgAdmin 4, etc).
* **T**ransform: what data cleaning or transformation was required.
* **L**oad: the final database, tables/collections, and why this was chosen.

**E**xtract: extracted data from three csv files found on Kaggle for Petfinder.com.

**T**ransform: We initially cleaned up our data by reading our files using pandas and creating a new dataframe by only selecting the columns we wanted to keep from the test.csv file. We then merged the submission.csv file with the test.csv file by merging on the “petid” column to add the Adoption Speed column that was missing from the test.csv file. We renamed columns to what made most sense to us. The original data used numbers to represent specific categories in some columns so we replaced those with the actual categories so that we won’t have to refer to an index to determine what the number stood for. We then filtered the type for “Dog” by using loc and separated our data for dogs and cats. Our data was mostly non-relational. We used the “petid” column as our primary key but the only two data we used it as foreign key were for the dog\_data and the cat\_data.

**L**oad: We used sqlachemy to load our final tables into pgAdmin. We had 11 final tables that we used to upload in our production database. We chose pgAdmin because it was more structured, and we know that we would have no missing data.