***Team members:***

Our team members are Fred Watts, Yuwen Dai, and Yuan Fang.

***Nomination for project coordinator:***

We nominate Fred Watts to be our project coordinator.

***Domain chosen***

Our domain is animal welfare. Millions of abandoned or lost animals end up in shelters. Many shelters are overcrowded, and some euthanize animals that aren’t adopted quickly. Studying adoption data may help shelters and rescue organizations connect with potential adopting homes more effectively, reducing needless animal suffering.

***Hypothesis or project topic***

An algorithm can be made to predict animal adoption rates from traits in the animal’s online profile. These traits may include breed, color, gender, health, the quality of images, and the quality of descriptive text.

***Available data sources (or where you’re going to get the data from)***

Our main data source is a dataset available at Kaggle.com on animal shelter records. Here are the tables we intend to use:

* *train.csv* - Tabular/text data for the training set
* *test.csv* - Tabular/text data for the test set
* *breed\_labels.csv –* Data dictionary on breed
* *color\_labels.csv* - Data dictionary on color
* *state\_labels.csv* - Data dictionary on location

If the capstone advisor would like to propose any further datasets, we are open to their suggestions.

***A brief description of what you will do***

* First, we will perform a data quality check and EDA on training and testing data, for example: analyzing missing values, summary statistics, data visualizations, scatter plots, and a correlation matrix. We may apply data manipulation depending on the quality of the data, for example: replacing missing data with mean, median, or other imputation techniques.
* Afterward, we will use machine learning techniques to narrow down the candidate variable list. Since our target variable is a categorical variable, it is a classification problem, so we will explore using classification approaches. The results of these prior steps will provide insight into the best format for our final model.
* Finally, we will build this model and further refine it through testing.
* If necessary, we will return to the first step and consider new traits to measure and model then repeat this process.

***Questions or avenues of exploration required for the project***

* What profile traits should we include in our model?
* Which traits have the most statistically useful records in the Kaggle tables?
* Considering these traits, which might suggest actions that could be taken to improve how animals are adopted?
* What may be the best classification model in our case? K - nearest neighbor? Decision trees? Support Vector machines?
* After our best model have been created, how can we utilize the model or present it in a way that could be easily used by shelters and rescue organizations?