Adoptions in Austin Animal Shelter – Review of Correlation Between Breed Attributes, Surrender Conditions

Team Ten – Chris Flint, Robert Perron, Cindy Gorbas

**April 14, 2019**

Approximately 68 % of American households have pets. Unfortunately, sometimes there are circumstances when pet owners are not longer able to keep their pets. We have decided to look at possible factors in successful dog adoptions. Using data from Dogtime.com and the Austin Animal Shelter, we sought to discover which factors could lead to successful adoption.

**Methodolgy**

**Extraction**

We decided we would limit our efforts to dogs…because we love dogs. From Kaggle.com we found two data sources that we believed would be helpful.

* From dogtime.com we retrieved a json file (rating.json). This json contained data such as breed, personality, health and activity. The data was ultimately read into a dataframe (dogs\_df). This data was read into a dataframe (shelter\_df)
* From the Austin Animal Center we used a csv file (aac\_intakes\_outcomes.csv). This file was extensive, and provided data such as reason for intake (admittance to the shelter), how long animal spent in the shelter, etc.

**Transform**

Once both sources of data were in dataframes, the data was transformed by:

Data Cleanup

* Renaming columns to eliminate wanted spaces in the column names  
  eliminating animal types (cat, other, etc) to ensure we were only looking at dogs
* Narrowing the data by removing Mixed breeds. Data from the Animal shelter included Mixed Breed data (ex. Labrador/husky). The data was not consistent and too varied from the Breed data from Dogtime.com (is a labrador/husky more husky or Labrador attributes?) .
* Parsing columns and removing unwanted characters.
* Removing NaN values.
* Dropping unwanted columns.

Addition of new, informative columns

* The percentage of good outcomes – returned to owner, adopted – were calculated and added to the dataset as *Good Outcome Percentage*.
* The means of the columns for the number of days in shelter and ages when arrived and adopted were calculated ad added to the dataset as*: time\_in\_shelter, outcome\_age\_days, intake\_age\_days*.
* The Breed attributes from the Dogtime.com data were aggregated into descriptive columns; these columns were added as *Friendly, Intelligence, Tolerance, Negative Attributes.* From these a new dataframe , dog\_attributes\_df, was created.
* The dataframes from different sources were merged into one dataframe.

The transformation of the data reduced the quantity from 7500 rows into 200 rows with data that is clean, useful and from of NaN.

**Load**

The team loaded the data using Mongo. Dataframes were converted into json files then loaded into mongo ‘database’.

* ETL\_project\_dogs – mongo database
* ‘dog\_data’ – mongo collection created by inserting/loading the final\_dog\_json (*final\_dog\_data\_df*) data
* ‘shelter\_data’ - mongo collection created by inserting/loading the final\_shelter\_json (*shelter\_details\_df*) data
* ‘total\_data’ - mongo collection created by inserting/loading the final\_data\_json (*final\_dog\_data\_df*) data
* ‘total\_data\_condensed’ - - mongo collection created by inserting/loading the condensed\_dog\_json (dog\_attributes\_df) data

**Resources**

**Data Sources:**

<https://www.kaggle.com/aaronschlegel/austin-animal-center-shelter-intakes-and-outcomes/version/1>

<https://www.kaggle.com/hocop1/cat-and-dog-breeds-parameters>

**Other Sources:**

<https://www.statista.com/statistics/198086/us-household-penetration-rates-for-pet-owning-since-2007/>