MARK5826 Product Analytics

Group 5

**Shelter Animal Outcomes: Progress Report**

Based on: [**https://www.kaggle.com/c/shelter-animal-outcomes/data**](https://www.kaggle.com/c/shelter-animal-outcomes/data)

**Part 1: The Founders**

Zhi He (Zachary) z5123906

I am in my last semester of master of information technology with four years’ business work experience, mainly responsible for the coding part in this project.

Strengths: Coding, problem solving, business advertising.

Urszula Adamczyk z3217453

Final year Masters of Commerce (Marketing) student. Media professional who has worked across a variety of fields, current Art Director at a creative agency.

Strengths: Ideation, problem solving, design, business strategy, marketing.

PeiGuo Guan z5143964

Master of information technology, majoring in data science field with four years’ e-commerce technology background. Responsible for the coding and data analysis.

Strengths: Coding, problem solving, data analyse.

Chuxin Huang z5152854

Part-time Master of Statistics and full-time reporting & insights specialist at an airline loyalty program. Responsible for business strategy, data analysis/visualisation and project presentation.

Strengths: project management, business strategy, data analysis and visualisation

**Part 2: About the project**

**Problem Statement:** Animals in shelters are at high risk of not being rehomed, and face the possibility of euthanasia.

**Project Aim:** Improving shelter animal outcomes and reducing the rate of euthanasia through identifying adoption trends and high risk individuals.

**The Problem**

Animals which enter shelters face an uncertain future, with many struggling to be adopted, and the threat of euthanasia incredibly high. While there is no legal requirement for pounds and shelters in Australia to report on the number of animals entering their facilities, it is estimated that in NSW alone, between 31-47% of dogs and 59-67% of cats will be euthanized in shelters (Getting To Zero, 2010). From research conducted by the Animal Welfare League of NSW, 90% of animals in shelters are savable (2015), however many still struggle to find families.

**The Solution**

Due to a lack of Australian data attributed to non-mandatory reporting (Chua, et al, 2017), our group will be basing its research on American data as a proof of concept, with the aim of bringing the findings to an Australian context. In America figures are significantly higher with 7.6 million animals entering shelters, 2.7 million of which are euthanized. The data set we have access to includes information about animal outcomes as well as their breed, color, sex, and age, from the Austin Animal Center between October 1st, 2013 to March, 2016 (Kaggle, 2016). Using this data we hope to create an accurate model that can identify trends in animal outcomes.

Through discovering trends in the types of animals struggling to be adopted we hope to help shelters be more tactical in the way they address the adoption of these animals. This will be done in multiple ways.

1. Incoming animals can be assessed based on the data and determine their risk level. High risk animals will be quickly identified when they enter the shelters. Staff will be able to give them extra care and extra promotion, to help these animals maximise their chances of adoption.
2. The trends can be used to inform adoption campaigns, for example if it discovers older dogs struggle to be adopted this can be used to in advertising to promote consideration of adopting older pets.
3. Trends can be integrated into a recommendation system on adoption websites, to suggest relevant high risk animals to individuals browsing the site.
4. Animals that have a high chance of adoption can be moved to pet stores, helping them rehome quickly by being placed in clear view of potential owners. This would also allow shelters to focus energy on the high risk animals.
5. Geographic data can be integrated in the future to determine if there are local trends. Using this data, animals can be promoted in different areas where they may have higher chances.

**Part 3: Plan**

**Table: Breakdown of tasks**

|  |  |
| --- | --- |
| **Tasks** | **Assigned to** |
| Creation of business plan | Urszula Adamczyk Chuxin Huang |
| Clean data | Zhi He  PeiGuo Guan |
| Descriptive analysis | Zhi He  PeiGuo Guan |
| Data visualisations | Urszula Adamczyk Chuxin Huang |
| Hypothesis development | All |
| Interpretation of data and Identification of trends | All |
| Writing final report | All |

**Table: Project Schedule**

|  |  |  |
| --- | --- | --- |
| **Week** | **Task** | **Assigned to** |
| 27th August | Progress report | All |
| 3rd September | Developing business plan and hypothesis development | Urszula Adamczyk  Chuxin Huang |
| 10th September | Data Cleaning and descriptive analysis | Zhi He  PeiGuo Guan |
| 17th September | Interpretation of data and identification of trends | Zhi He  PeiGuo Guan |
| 24th September | Data visualisation and draft report | Urszula Adamczyk  Chuxin Huang |
| 1st October | Writing report. Finalising the business plan and report findings | All |
| 5th October | Due | All |

**Methodology**

In order to predict the outcome of the animal as they leave the animal center, these outcomes include: adoption, died, euthanasia, return to owner, and Transfer, we utilize several tools to establish the machine learning environment. The following tools are used:

* Tableau Desktop 2018.1

We utilize tableau to analyze data and create data visualizations, publish data sources as well as workbooks to tableau Server

* Python Scripting Environment   
  We choose python as the programming language to implement the machine learning algorithm, train our model, and test the hypothesis.
* Kaggle

We got the original prototype of this project from kaggle, kaggle machine   
           learning competition inspire us to choose this project as our group work, and      
           kaggle kernels also provide a ideal cloud based work bench for our data     
           science analyst.

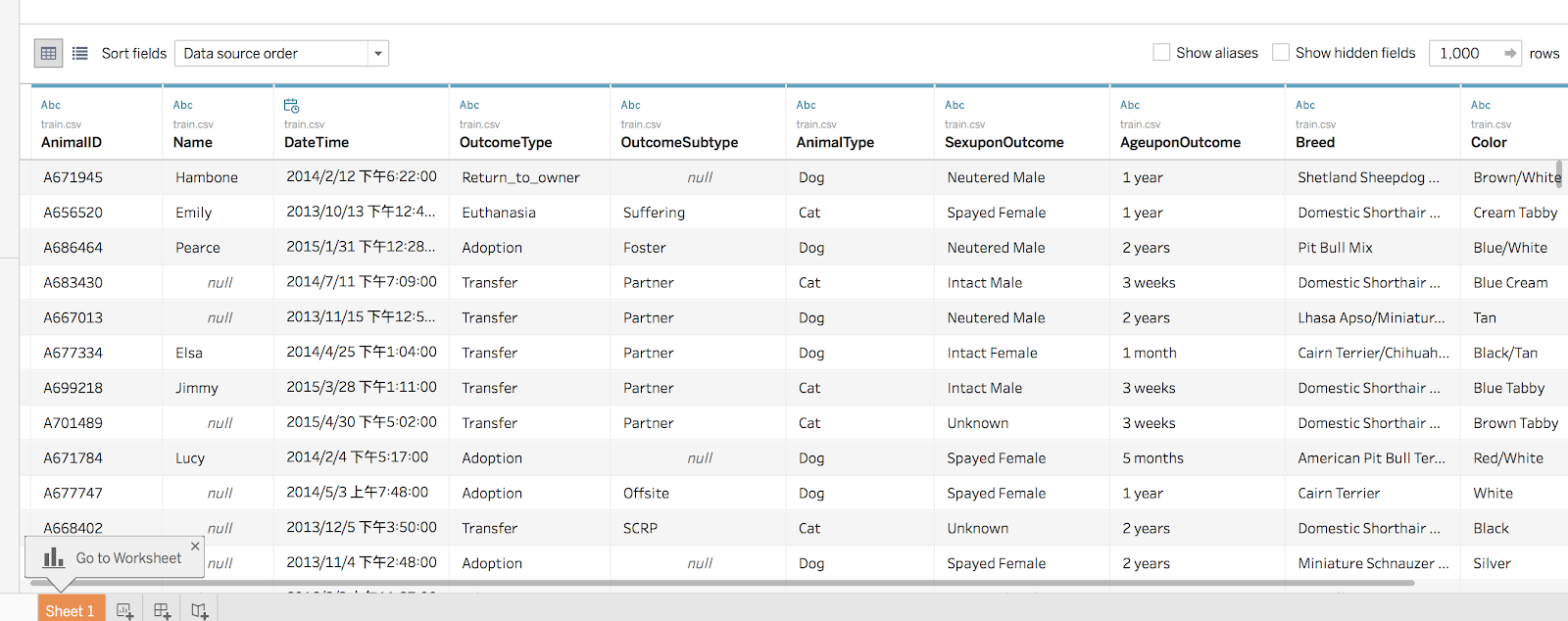
Data source

The data comes from [Austin Animal Center](http://www.austintexas.gov/department/animal-services) from October 1st, 2013 to March, 2016. Outcomes represent the status of animals as they leave the Animal Center. All animals receive a unique Animal ID during intake.

File descriptions

* train.csv - the training set
* test.csv - the test set

The train and test data are randomly split. The following table shows the structure of data.



We will use machine learning to train data and make a data model based on python skills. After the process of data processing like cleaning, classification and so on, we will use logistic regression to analyses and record our data at the first stage, which is followed by the probability of the strategy change according to the performance of result. Based on the result of data, the data will be presented by tableau which will make result and prediction more visualiable and understandable.

**References**

Animal Welfare League of NSW, 2015, *‘Canine And Feline Statistics’*, Viewed 28 August 2018, <<http://www.awlnsw.com.au/statistics.html>>.

Australian Veterinary Association, 2016, *‘Pets, owners and the rise of the fur baby’*, Viewed 28 August 2018,

<<https://www.vetvoice.com.au/ec/pet-ownership/pet-ownership-statistics/>>.

Chua, D., Rand, J., & Morton, J., 2017, ‘Surrendered and Stray Dogs in Australia-Estimation of Numbers Entering Municipal Pounds, Shelters and Rescue Groups and Their Outcomes’, *Animals : An Open Access Journal from MDPI*, *7*(7), 50, Viewed 28 August 2018, <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5532565/>>.

Getting To Zero, 2010, *‘Estimated Number of dogs and cats abandoned and killed in Australia 2009/10’*, Viewed 28 August 2018,

<<https://www.g2z.org.au/pdf/Calculation%20of%20National%20Figures%2009%2010%20incl.%20Healthy%20Treatable.pdf>>

Kaggle, 2016, *‘Shelter Animal Outcomes’*, Viewed 28 August 2018, <<https://www.kaggle.com/c/shelter-animal-outcomes>>.