**Name:** Nguyen Thi Cam Hoang **Date:** Nov 28th, 2016

**Student ID:** 20165327

**Brief report**

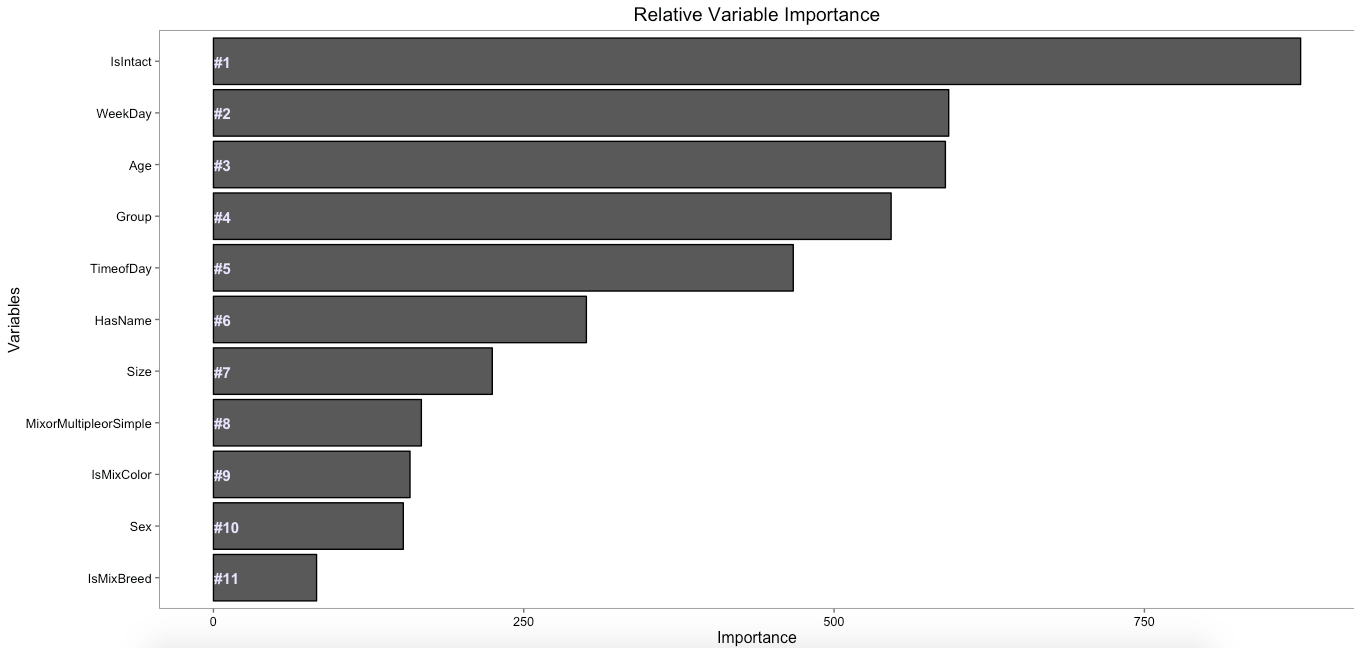
**Data preprocessing**

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **Detail** | **Type** | **TODO** |
| **Dependent variable** | | | |
| OutcomeType | Adoption/Died/Euthanasia  Return\_to\_owner/Transfer | Categorical | LabelEncoder |
| **Independent variables** | | | |
| HasName | 0/1 | Categorical |  |
| TimeofDay | morning, midday, lateday, night | Categorical |  |
| WeekDay | 0, 1, 2, 3, 4, 5, 6 | Categorical |  |
| IsIntact | 0/1 | Categorical |  |
| Sex | Male/Female | Categorical |  |
| Age | Juvenile, Adult, Old 🡪 0, 1, 2 | Ordinal | Mapping |
| IsMixBreed | 0/1 | Categorical |  |
| MixorMultipleorSimple | Mix, Multiple, Simple | Categorical |  |
| Group | Sporting, NonSporting, Hound, Working, Terrier, Toy, Herding, Pitbull, Unknown | Categorical |  |
| Size | S/M/L/G 🡪 0, 1, 2, 3 | Ordinal | Mapping |
| IsMixColor | 0/1 | Categorical |  |

**Models**

## Random Forest

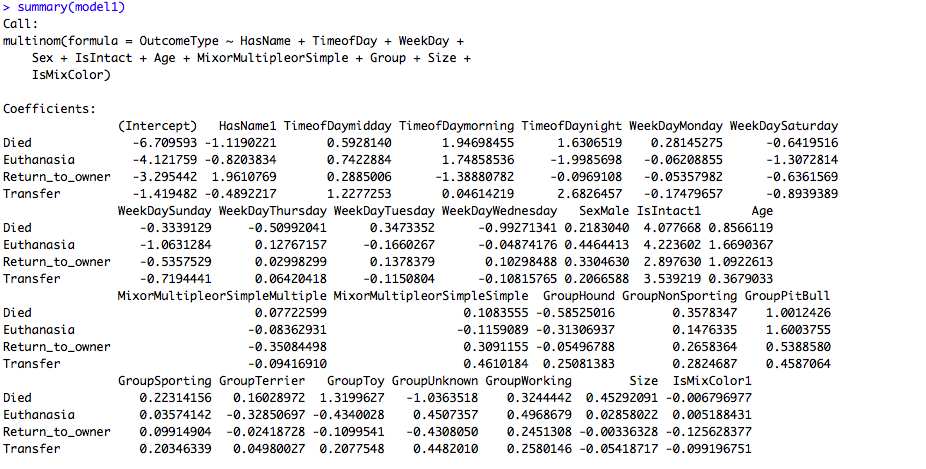
* **Note**: This model is run in R

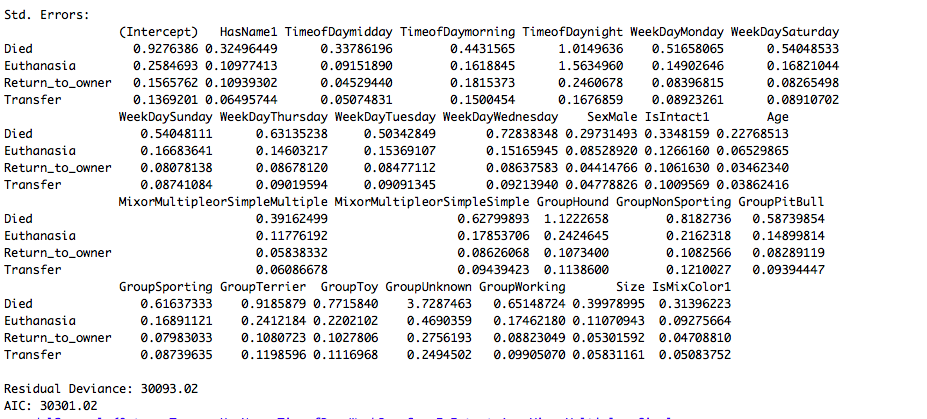


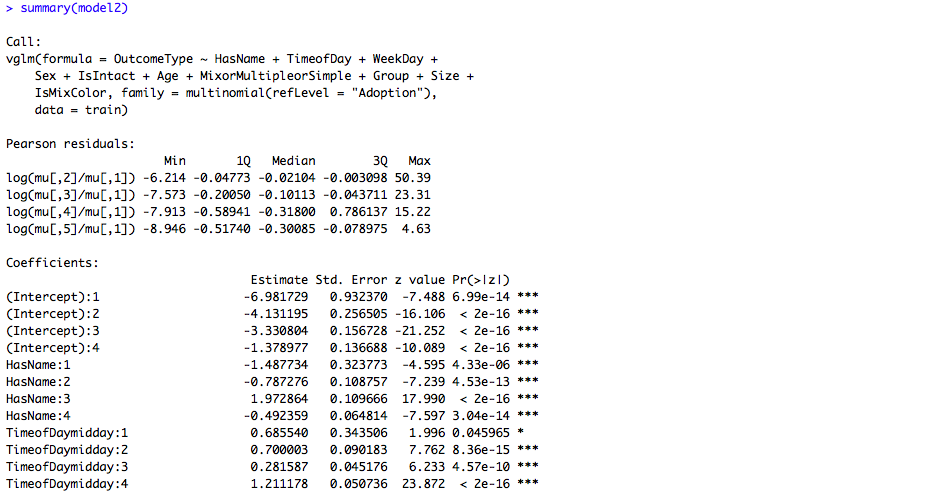
* **Note**: This model is run in Python
* Choose variables based on RandomForest: Variable **IsMixBreed** and **MixorMultipleorSimple** is highly correlated. Based on the results from RandomForest, MixorMultipleorSimple seems to be more important. Therefore, IsMixBreed is eliminated from the model.
* From now on, use all variables *except* **MixorMultipleorSimple** for all models.
* 10 independent variables
* 32 levels

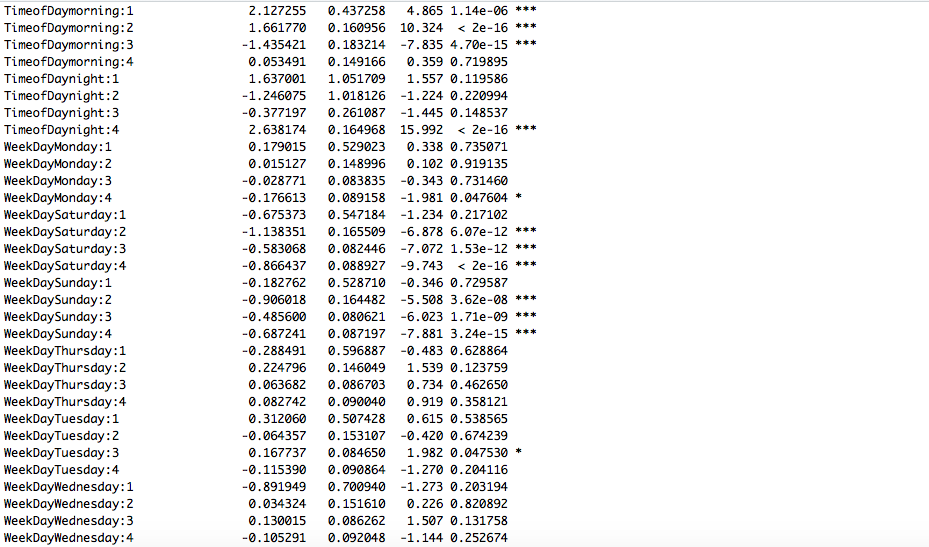
## Multinomial regression

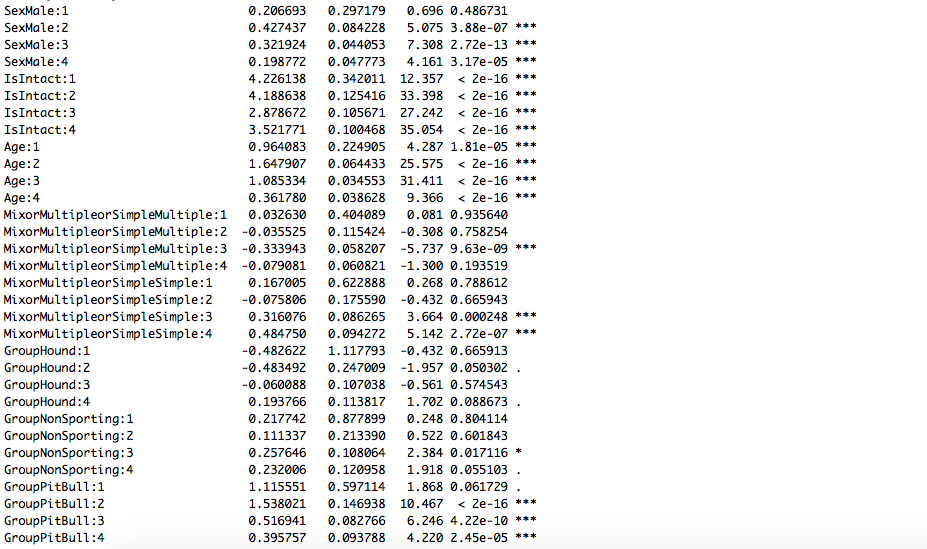
* **Note**: This model is run in R and Python
* Multinomial regression:

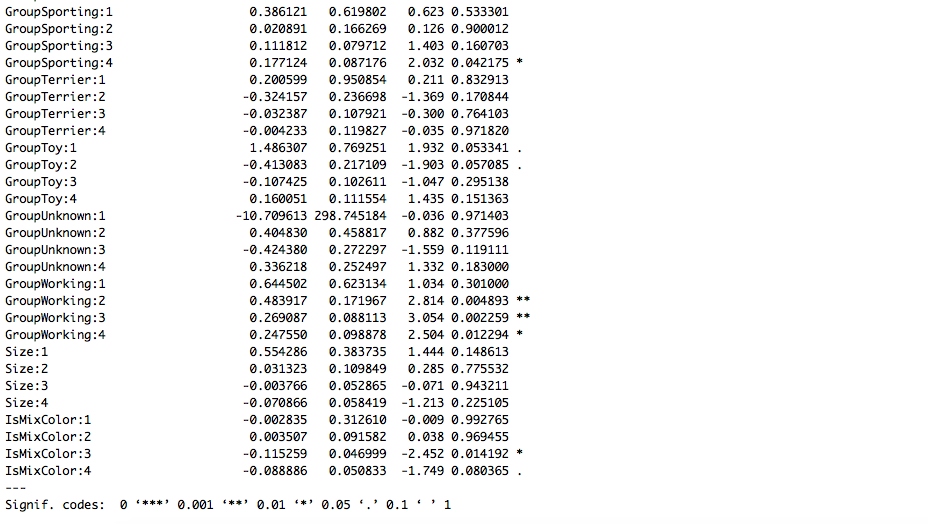


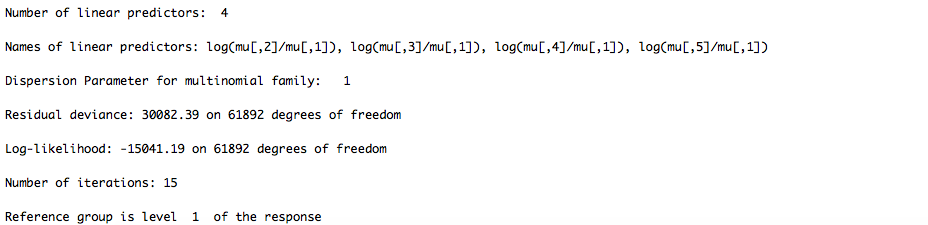


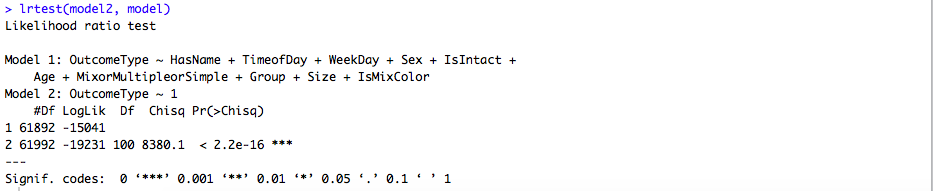










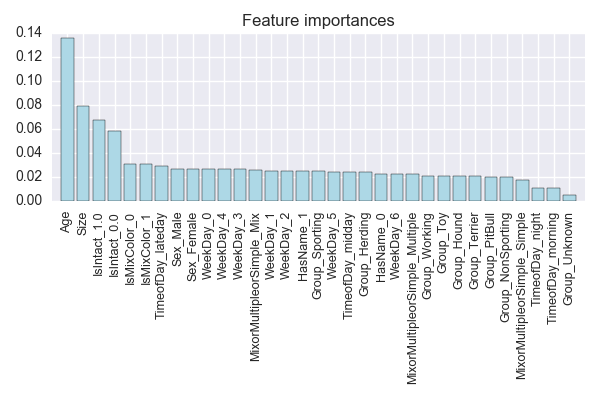


|  |  |
| --- | --- |
|  |  |
|  |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Variable name and levels** | **Odds ratio** | | | | |
| **Adoption** | **Died** | **Euthanasia** | **Return\_to\_owner** | **Transfer** |
| Intercept | 1.21763 | -2.86396 | -0.74539 | -0.08372 | 2.47544 |
| Age | -0.80330 | 0.15031 | 0.82237 | 0.30787 | -0.47725 |
| Size | -0.03467 | 0.13429 | 0.02511 | -0.02605 | -0.09868 |
| HasName\_0 | -0.09170 | 0.72873 | 0.30464 | -1.09777 | 0.15610 |
| HasName\_1 | 0.09170 | -0.72873 | -0.30464 | 1.09777 | -0.15610 |
| TimeofDay\_lateday | 0.37092 | -0.55515 | -0.00091 | 0.78534 | -0.60021 |
| TimeofDay\_midday | -0.22345 | -0.30321 | 0.03568 | 0.46049 | 0.03048 |
| TimeofDay\_morning | 0.01771 | 0.49718 | 1.32754 | -0.96211 | -0.88033 |
| TimeofDay\_night | -0.16518 | 0.36117 | -1.36231 | -0.28373 | 1.45005 |
| WeekDay\_0 | -0.18967 | 0.31183 | 0.06574 | -0.10302 | -0.08488 |
| WeekDay\_1 | -0.20680 | 0.16978 | -0.00682 | 0.08111 | -0.03726 |
| WeekDay\_2 | -0.07194 | -0.49322 | 0.25637 | 0.20802 | 0.10077 |
| WeekDay\_3 | -0.15613 | -0.25818 | 0.26951 | -0.03323 | 0.17803 |
| WeekDay\_4 | -0.16946 | 0.05892 | 0.08430 | -0.08499 | 0.11123 |
| WeekDay\_5 | 0.48055 | 0.03489 | -0.40990 | 0.01549 | -0.12102 |
| WeekDay\_6 | 0.31347 | 0.17598 | -0.25919 | -0.08338 | -0.14687 |
| IsIntact\_0.0 | 1.48728 | -0.61385 | -0.61030 | 0.02750 | -0.29062 |
| IsIntact\_1.0 | -1.48728 | 0.61385 | 0.61030 | -0.02750 | 0.29062 |
| Sex\_Female | 0.10932 | 0.00991 | -0.09521 | -0.05336 | 0.02934 |
| Sex\_Male | -0.10932 | -0.00991 | 0.09521 | 0.05336 | -0.02934 |
| MixorMultipleorSimple\_Mix | 0.01164 | -0.03485 | 0.08532 | 0.04132 | -0.10343 |
| MixorMultipleorSimple\_Multiple | 0.15693 | 0.00063 | 0.12113 | -0.18437 | -0.09432 |
| MixorMultipleorSimple\_Simple | -0.16856 | 0.03422 | -0.20645 | 0.14304 | 0.19775 |
| Group\_Herding | 0.15576 | -0.19783 | -0.01425 | 0.08213 | -0.02581 |
| Group\_Hound | 0.24556 | -0.17157 | -0.37356 | 0.08691 | 0.21266 |
| Group\_NonSporting | 0.00685 | -0.25127 | 0.00093 | 0.22315 | 0.02034 |
| Group\_PitBull | -0.57384 | 0.34153 | 0.82142 | -0.19210 | -0.39701 |
| Group\_Sporting | -0.00955 | 0.20360 | -0.27157 | 0.10254 | -0.02502 |
| Group\_Terrier | 0.37807 | -0.69655 | -0.07360 | 0.25609 | 0.13599 |
| Group\_Toy | 0.01665 | 0.75383 | -0.54551 | -0.17897 | -0.04599 |
| Group\_Unknown | 0.00242 | -0.13629 | 0.25678 | -0.39497 | 0.27206 |
| Group\_Working | -0.22192 | 0.15455 | 0.19937 | 0.01522 | -0.14721 |
| IsMixColor\_0 | -0.03386 | 0.01078 | -0.02340 | 0.01612 | 0.03037 |
| IsMixColor\_1 | 0.03386 | -0.01078 | 0.02340 | -0.01612 | -0.03037 |

## Random Forest

* **Note**: This model is run in Python



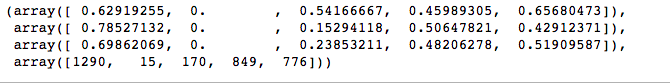
* Accuracy = 0.55
* Frequency:
* Adoption:0.65
* Died: 0
* Euthanasia: 0.20
* Return\_to\_owner: 0.43
* Transfer: 0.49

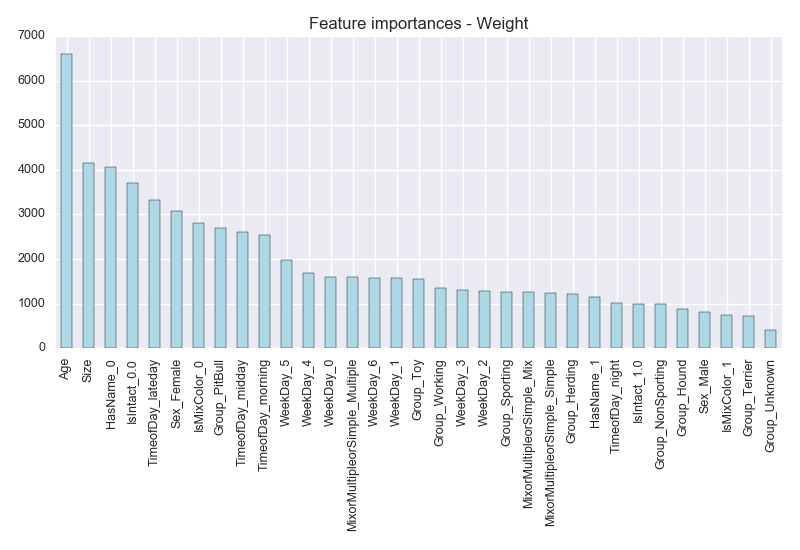
### Interpretation

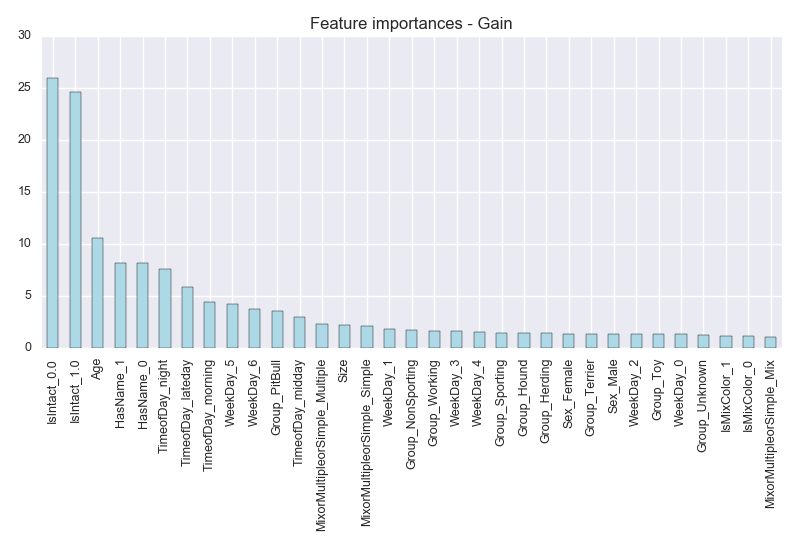
* Model explains well the outcome of adoption, return\_to\_owner and transfer.
* For the other two variables, the model cannot predict because the lack of observation.

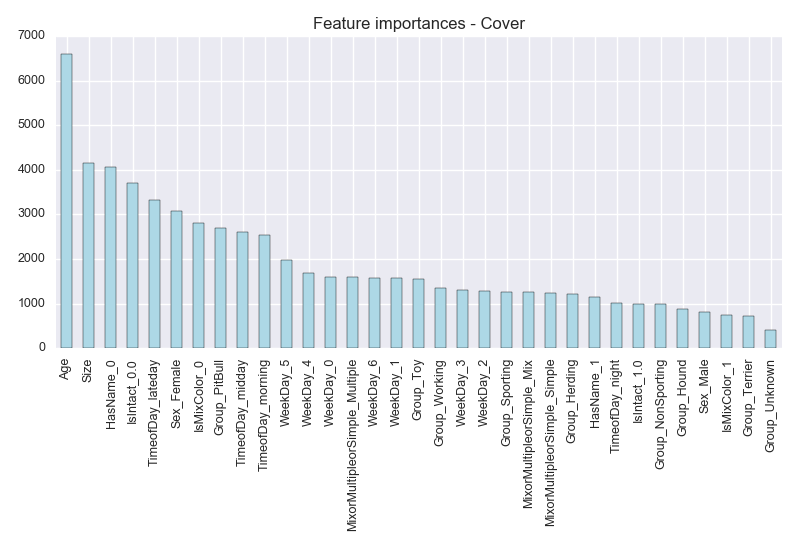
## XGBoost

* **Note**: This model is run in Python
* Accuracy = 58%





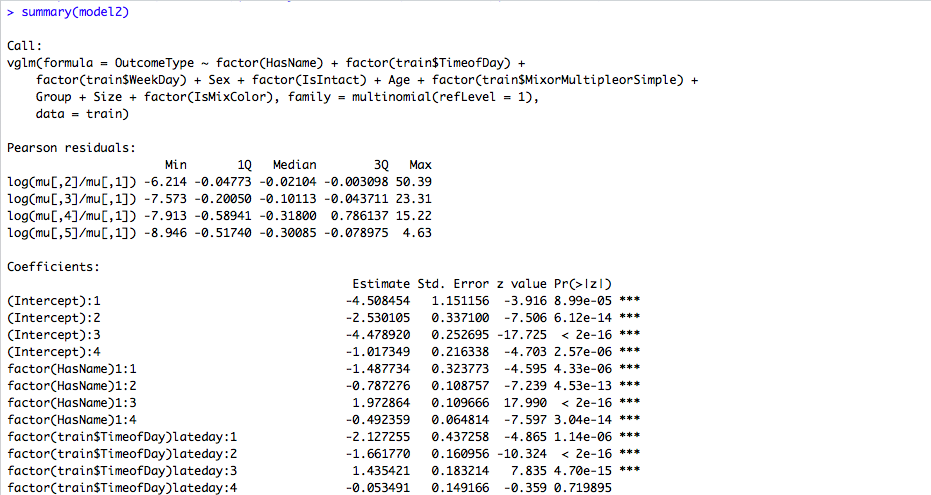


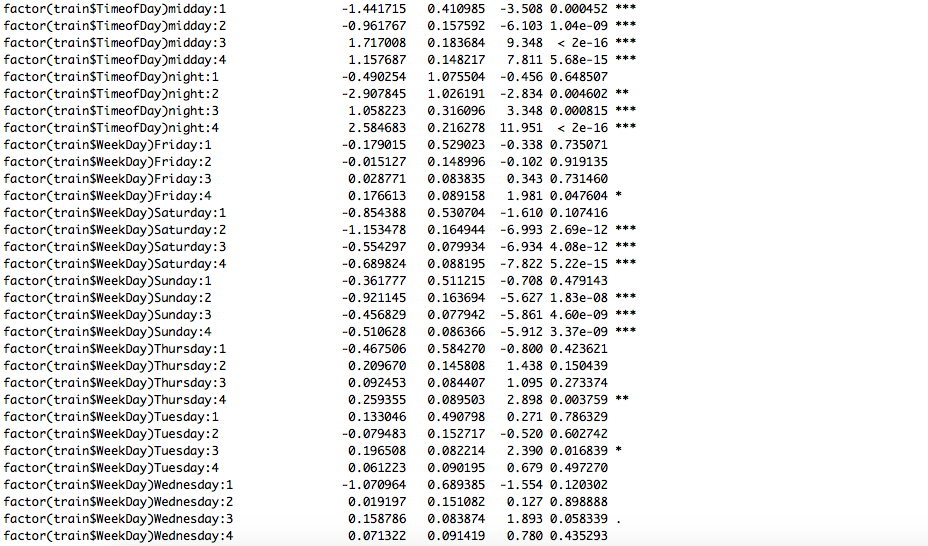


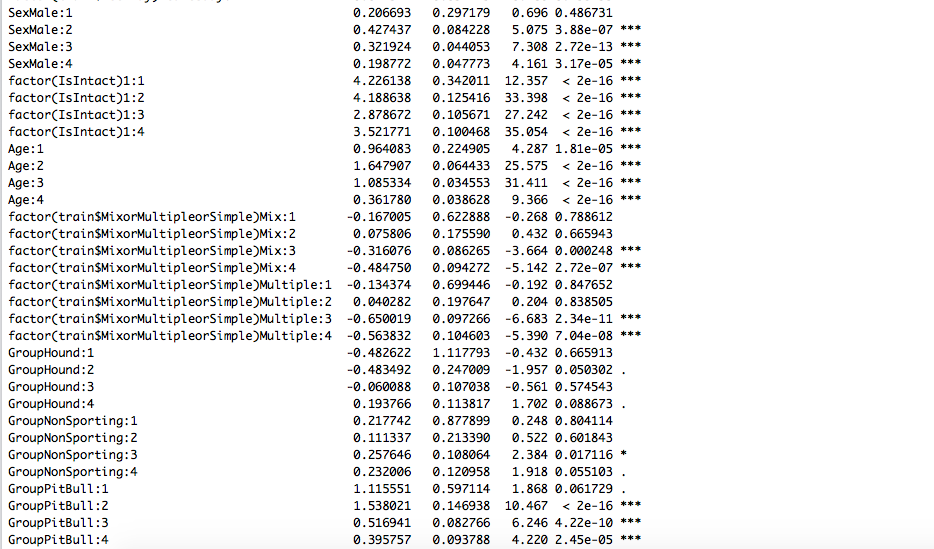
1. The Gain implies the relative contribution of the corresponding feature to the model calculated by taking each feature's contribution for each tree in the model. A higher value of this metric when compared to another feature implies it is more important for generating a prediction.
2. The Cover metric means the relative number of observations related to this feature. For example, if you have 100 observations, 4 features and 3 trees, and suppose feature1 is used to decide the leaf node for 10, 5, and 2 observations in tree1, tree2 and tree3 respectively; then the metric will count cover for this feature as 10+5+2 = 17 observations. This will be calculated for all the 4 features and the cover will be 17 expressed as a percentage for all features' cover metrics.
3. The Frequence (frequency) is the percentage representing the relative number of times a particular feature occurs in the trees of the model. In the above example, if feature1 occurred in 2 splits, 1 split and 3 splits in each of tree1, tree2 and tree3; then the weightage for feature1 will be 2+1+3 = 6. The frequency for feature1 is calculated as its percentage weight over weights of all features.

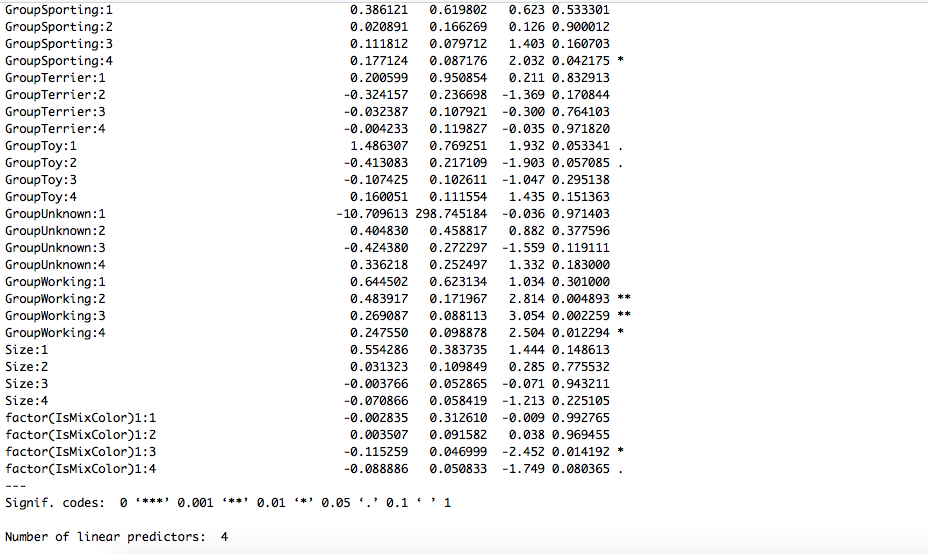
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Variable name and levels** | **Odds ratio** | | | | |
| **Adoption** | **Died** | **Euthanasia** | **Return\_to\_owner** | **Transfer** |
| Intercept | 1.1801698 | -2.80256871 | -0.72905616 | -0.04801812 | 2.39947319 |
| Age | -0.80591156 | 0.13690816 | 0.83547662 | 0.27842438 | -0.4448976 |
| Size | -0.07290339 | 0.32291131 | -0.03070819 | -0.07590826 | -0.14339146 |
| HasName\_0 | -0.07325505 | 0.63749166 | 0.31773328 | -1.05450778 | 0.17253789 |
| HasName\_1 | 0.07325491 | -0.63749319 | -0.31773125 | 1.05450879 | -0.17253926 |
| TimeofDay\_lateday | 0.41675203 | -0.59984404 | -0.05091379 | 0.79087529 | -0.55686949 |
| TimeofDay\_midday | -0.15740431 | -0.49389236 | 0.07397677 | 0.49844015 | 0.07887974 |
| TimeofDay\_morning | -0.03639296 | 0.92172918 | 1.13838044 | -1.07260977 | -0.95110689 |
| TimeofDay\_night | -0.22295489 | 0.17200569 | -1.16144139 | -0.21670466 | 1.42909526 |
| WeekDay\_0 | -0.15728206 | 0.1905389 | 0.11503093 | -0.08104193 | -0.06724584 |
| WeekDay\_1 | -0.22317624 | 0.27326211 | -0.0270499 | 0.04899633 | -0.07203231 |
| WeekDay\_2 | -0.03505861 | -0.55051448 | 0.25958874 | 0.1994841 | 0.12650025 |
| WeekDay\_3 | -0.19256136 | -0.23153089 | 0.29221808 | -0.0242174 | 0.15609157 |
| WeekDay\_4 | -0.16552139 | 0.02994081 | 0.09547344 | -0.06037324 | 0.10048038 |
| WeekDay\_5 | 0.48148046 | 0.01737956 | -0.38718638 | 0.00462205 | -0.1162957 |
| WeekDay\_6 | 0.29211906 | 0.27092247 | -0.34807289 | -0.0874689 | -0.12749973 |
| IsIntact\_0.0 | 1.47220521 | -0.61725502 | -0.61275828 | 0.03975682 | -0.28194873 |
| IsIntact\_1.0 | -1.47220534 | 0.61725349 | 0.6127603 | -0.0397558 | 0.28194735 |
| Sex\_Female | 0.11498418 | 0.01288864 | -0.09791585 | -0.04575457 | 0.0157976 |
| Sex\_Male | -0.11498432 | -0.01289017 | 0.09791787 | 0.04575558 | -0.01579897 |
| MixorMultipleorSimple\_Mix | 0.02941479 | -0.02613354 | 0.06641418 | 0.03553327 | -0.10522871 |
| MixorMultipleorSimple\_Multiple | 0.12069394 | 0.06004847 | 0.11916014 | -0.20676003 | -0.09314253 |
| MixorMultipleorSimple\_Simple | -0.15010887 | -0.03391647 | -0.1855723 | 0.17122778 | 0.19836987 |
| Group\_Herding | 0.1471601 | -0.18353214 | -0.00444256 | 0.08572029 | -0.0449057 |
| Group\_Hound | 0.27468738 | -0.36976383 | -0.33530095 | 0.15404077 | 0.27633662 |
| Group\_NonSporting | -0.01123797 | -0.160007 | -0.04163943 | 0.18468358 | 0.02820081 |
| Group\_PitBull | -0.5365071 | 0.10594142 | 0.84620782 | -0.0816835 | -0.33395864 |
| Group\_Sporting | 0.02678967 | -0.01684477 | -0.09868447 | 0.07726722 | 0.01147235 |
| Group\_Terrier | 0.20941728 | -0.09350974 | -0.24662367 | 0.11679378 | 0.01392235 |
| Group\_Toy | -0.02026424 | 0.82549098 | -0.56504409 | -0.18762796 | -0.05255469 |
| Group\_Unknown | 0.08340706 | -0.21229967 | 0.29167955 | -0.38296006 | 0.22017312 |
| Group\_Working | -0.17345232 | 0.10452321 | 0.15384983 | 0.03376689 | -0.11868761 |
| IsMixColor\_0 | -0.02159708 | -0.0129665 | -0.02422091 | 0.03592885 | 0.02285564 |
| IsMixColor\_1 | 0.02159694 | 0.01296497 | 0.02422294 | -0.03592784 | -0.02285701 |

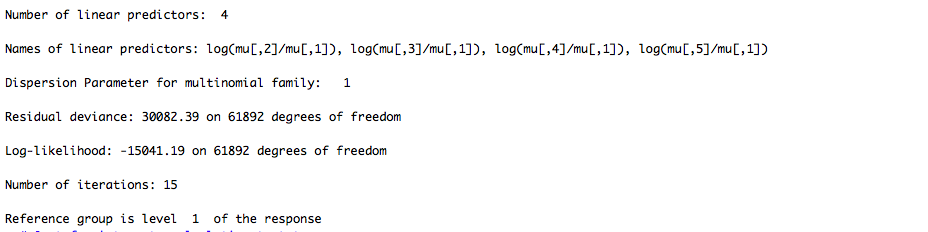
Model when set reference:

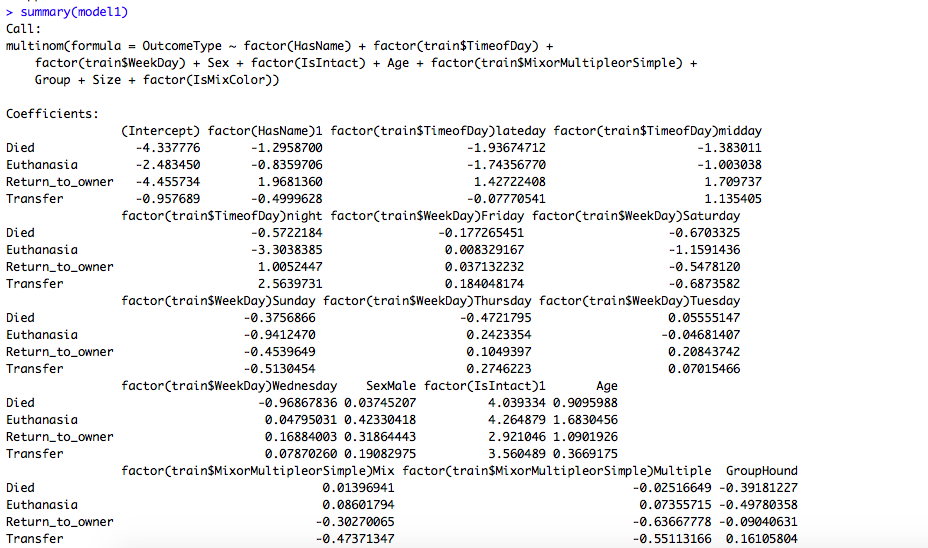


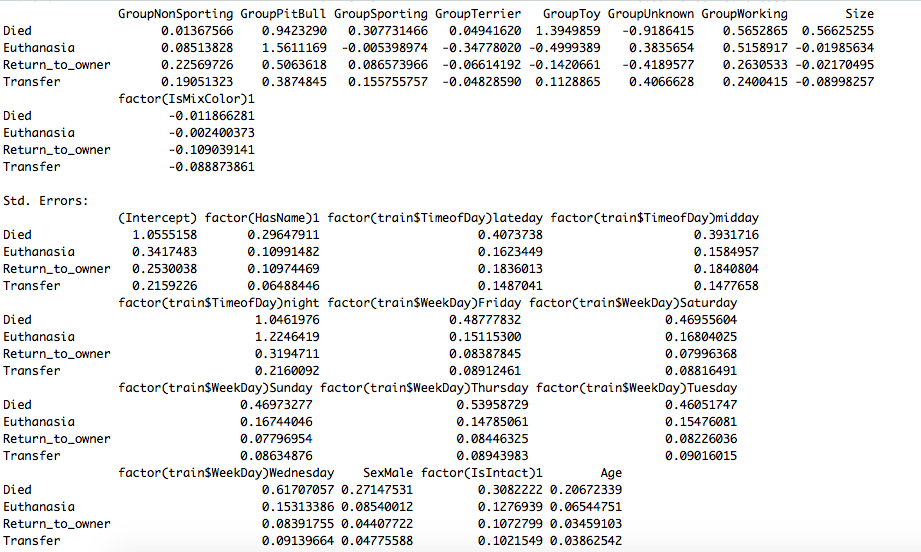


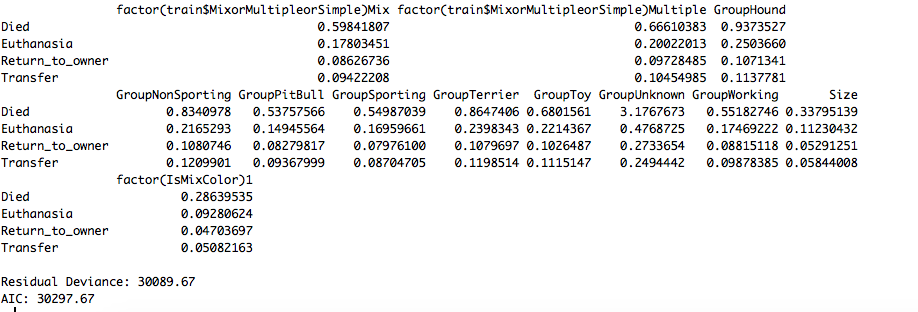




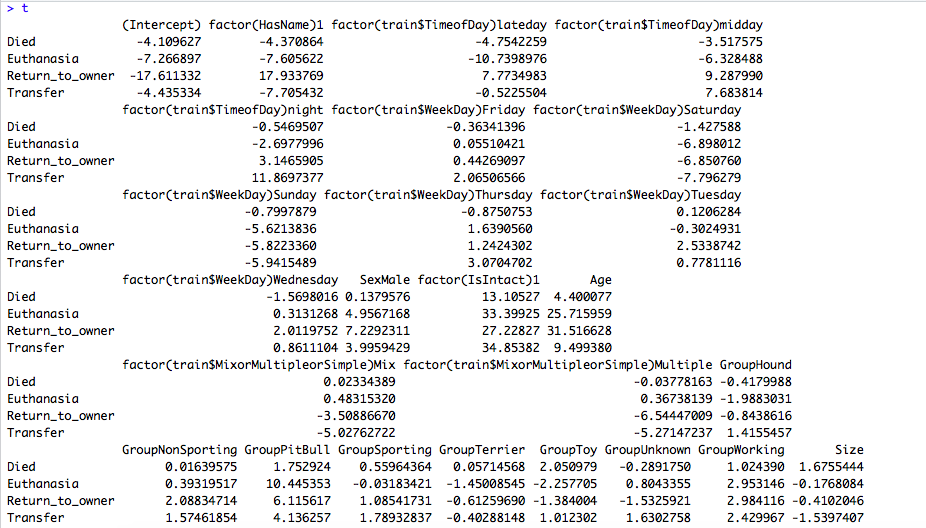








t stats





R = 0.2178774