Makenzie Barber

18 July 2022

Conditionals

Shelters get a bad reputation from the public as inhumane due to their population control efforts. In this report, I will disprove that reputation through multiple data analysis. We also will be disproving that animals of one species or another are euthanized more often because of their species alone. Calculating the number of animals in this report is key for shelters to know how many of each animal they have on hand. Keeping track of outcomes for cats and dogs alike can aid a shelter in preventing death in young animals, tracking euthanasia for population purposes, and to celebrate all adoptions and returns to owners. Since the dataset is spanned over three years, ordering what time an animal’s outcome was determined is key for obtaining records. This data was easy to work with given the lack of data integrity issues and plethora of data available. In the future, this data can be useful compared to data for shelters during the COVID-19 pandemic, as more people were in search of a companion than ever while confined to the walls of their homes. To further my initial statement

Chart, bar chart

Description automatically generated In the dataset from “shelter.csv”, we are provided with over 21,000 rows across 10 variables. In this part of the report, I will breakdown the data to better comprehend what is in it. This data comes from a Knoxville shelter over the span of 3 years (2013-2016) and reports the outcome for a specific animal, their name, animal ID, the full data and time at which their outcome was determined. To begin the investigation into this data, I separated the number of dogs from cats (figure 1), and discovered there were significantly more dogs than cats in this shelter (12,641 to 9,088). For the vectors of class “character”, I reclassified them as factors with levels for ease of use. For example, “OutcomeType” is now a factor with 5 levels, and “AgeuponOutcome” is a factor with 45 levels.

To demonstrate the type of outcome faced towards all animals listed, Figure 2 was developed to display the number of animals correlated with their outcome. However, in Figure 3, a mosaic plot was developed to display more detail in the difference in outcomes for dogs vs. cats. This plot was used with the associate function to compare 2 categorical variables of class factor. This compared the species of animal with the outcome it faced, and with a p-value of 0 and 95% confidence, I can confidently reject the null hypothesis that the type of animal is not correlated with any specific outcome type. However, this comparison did prove that cats do face euthanasia more than dogs in this specific shelter. Next, we will move to conditional functions that help define specific variables for comparison.

The conditional functions provided to us aid in isolating variables within a column and formatting them so that a user conditionally filter through variables and select only the levels they wish to display. In Figure 4, I used the “ifelse” statement to name a “Happy Ending” as Outcome Type that ended in Adoption. I nested another ifelse into the original statement to label a “Sad Ending” as an Outcome Type that ended in Death or Euthanasia. For the “Still a Happy Ending”, I nested yet another ifelse statement to only apply to those Outcome Types that Ending in a Transfer or a Return to owner. This graph is similar to Figure 2, except it is more straightforward and tells the user exactly what sort of outcome an animal received. Next, I will use the “lubridate” package to format and order the time represented in this dataframe. Chart, bar chart

Description automatically generated

Figure

In order to keep order the data, the user must utilize the “lubridate” package. With this set of data being mostly categorical, the only column it would make sense to use this package on is the “DateTime” column. To begin, I formatted the aforementioned column as a ymd\_hms (Year, Month, Day, Hour, Minute, Second) based on the data provided. Then, I used the “order()” function to order all rows by the date provided. This is important for shelters to keep records of where each animal goes if it leaves the shelter, in one way or another.

Figure

In this data, I have found that the “Names” column was missing some rows, however, the data was of upmost completeness. This could be from an overpopulation of animals, and therefore a data entry issue. No further integrity issues were discovered.

For future research, I would compare this data to data from the shelter over the span of the COVID-19 pandemic in the United States across the same shelter, as it was almost the same amount of time. This could be compared as more people searched for companions while isolating. This could also be used to urge the public to spay/neuter their animals unless they intentionally plan to breed them, given the number of deaths and overpopulation within animal shelters. There are many ways to use this data in the future, and since there is a lot of data, the possibilities are endless.

Figure 3
 Overall, this dataframe is among the largest I have ever worked with, and there is much to unpack. After converting the Animal Type to a factor, I was able to conclude that dogs have a much higher outcome of shelter-life, but cats unfortunately face euthanasia/death more often. Conditional statements allowed me to filter this data through ifelse statements, and allowed for shorter titles of outcomes, resulting in a much more aesthetically pleasing graph. The graphical representation revealed that more animals have a happy ending than a sad in shelters, and I believe that is a good outcome in any case. This data set lacked integrity issues, however, it had an excessive number of rows, which made it harder to work with and filter through. To conclude, this report revealed that although shelters face a massive overpopulation problem, they are not inhumane, as most animals receive a happy ending, no matter the kind of animal they are.

Chart, bar chart

Description automatically generated

Figure [4]