Angie Oehler
DSC680
Project Milestone 3

Dogs! (Or Finding a New Best Friend)

Introduction. This paper explains the process for finding a new dog for a person or family. The elements discussed will include what the problem is; background of the data and data dictionary; methods and analysis; and conclusions drawn. Assumptions about the data and it's limitations, as well as challenges will also be considered. Then, the paper will move on to recommendations and potential implementation plans. Finally, any ethical issues will be examined.

Business Problem. The problem this paper looks to solve is how to match a person with the best new dog for them. This could be for a family with small kids, an active single person, or a senior looking for companionship. Each of these groups have different needs and abilities for caring and/or training a dog. Helping people find the right dog helps to manage expectations, which will foster better relationships with their new pet.

Background/History. Looking for a dog can be exciting; however, we want a pet that will fit into our life without too much anxiety for us or the animal. Going to a new home full of strangers can be a stressful event for a dog. This can be reduced if we plan ahead and know what we are looking for and and know what to expect from various breeds of dogs.

Data Explanation. This project combines two separate datasets. The first is information provided by the American Kennel Club and contains the following data points (Fishman):

- The breed name
- The lower and upper height for the breed
- The upper and lower weight for the breed

The second dataset provides information collected by Dr. Stanley Coren, a dog behavior specialist at the University of British Columbia. This dataset gives data about:

- The breed name
- The AKC size and breed classification (Aerts):
  - Brightest Dogs - understands new commands in five or fewer repetitions and obeys first commands greater than or equal to 95% of the time.

Angie Oehler
DSC680
Project Milestone 3

- Excellent Working Dogs - understands new commands within 5 to 15 repetitions and obeys first commands greater than or equal to 85% of the time.
  Above Average Working Dogs - understands new commands within 15 to 25 repetitions and obeys first command greater than or equal to 70% of the time.
- Average Working/Obedience Intelligence - understands new commands within 25 to 40 repetitions and obeys first commands greater than or equal to 50% of the time.
- Fair Working/Obedience Intelligence - understands new commands within 40 to 80 repetitions and obeys first commands greater than or equal to 30% of the time.
- Lowest Degree of Working/Obedience Intelligence - understands new commands with an 80 to 100 repetitions and obey his first commands 25% of the time or less.

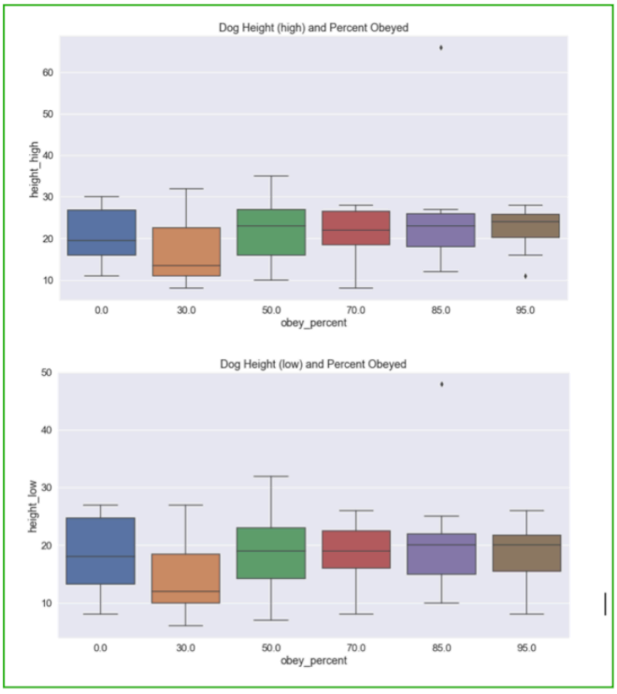
- The likelihood the breed will obey initial commands
- The lower and upper repetition limits for each bread to understand newly given commands

Methods. To begin analysis on these two data sets, the breed names were standardized according to the American Kennel Club's naming convention and then combined into a single data frame. The data was examined in a number of different ways, which will be discussed in the next section. Dog breeds without intelligence classifications were removed from the dataset because they could not be categorized.

Analysis. I began by a general exploration of the data, which included counts, min/max/ mean, percentile, and standard deviation for each column. I also looked at the general distributions of the data and did multiple comparisons to see how the variables stacked up against each other.

Box-plots (figure 1) proved very informative

Figure 1
Boxplots:
Comparison between Dog Heights and Percentages Obeyed



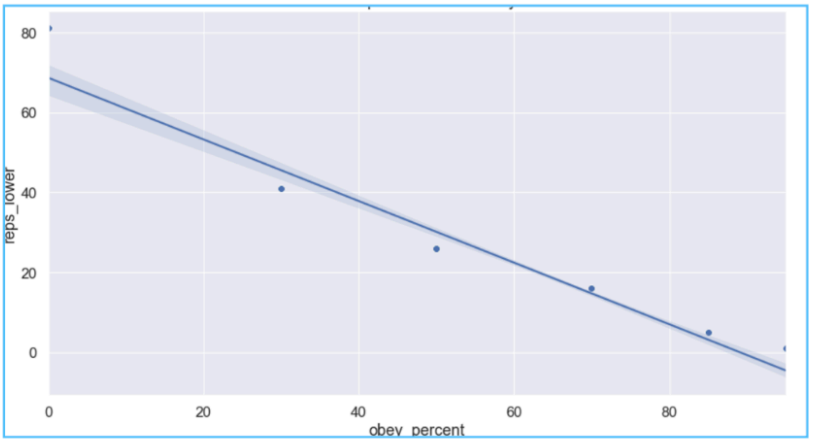
when looking at the number of repetitions and the percentages the dogs obeyed first commands. A side-by-side comparison of of the minimum/maximum scores and the distributions within each percentile provided the opportunity to see the averages for each grouping of dog heights, both high and low.

Knowing the distribution for the dog intelligence classifications was also important. The total counts for each class are as follows:

- Average Working/Obedience - 38
- Above Average Working Dogs - 27
- Fair Working/Obedience - 20
- Excellent Working Dogs - 19
- Lowest Working/Obedience - 10
- Brightest Dogs - 10

Figure 2

Bivariate Plot:
Lower Rep Counts and Percentages Obeyed



Next, I looked at the relationship between rep counts and percentages obeyed. With a bivariate scatterplot (figure 2), I can see that there is a strong relationship between the lower reps needed for new commands and a high percentage rate for first commands. As can be seen in the shaded area, there is more variance between the number of repetitions needed and the percent obeyed as the rep number increases. Once the number of repetitions is decreased and the percent of first commands obeyed increases, the relationship strengthens.

Figure 3

Dog Classifications:
Comparison of Average Lowest reps and Average Percentages Obeyed

Classification	obey_percent
Above Average Working Dogs	70.0
Average Working/Obedience Intelligence	50.0
Brightest Dogs	95.0
Excellent Working Dogs	85.0
Fair Working/Obedience Intelligence	30.0
Lowest Working/Obedience Intelligence	0.0

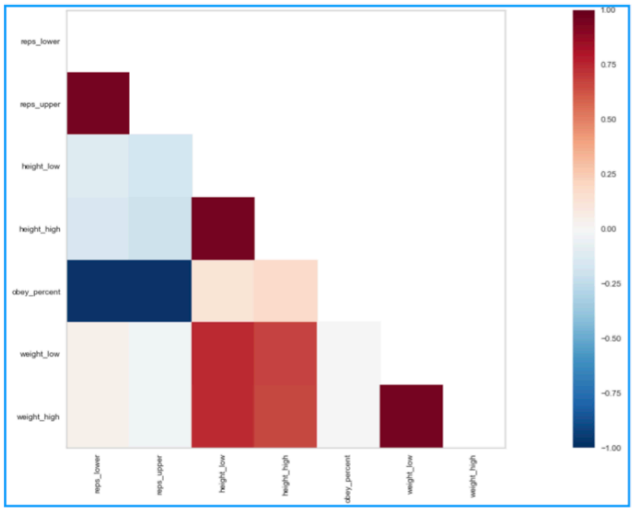
Classification	reps_lower
Above Average Working Dogs	16.0
Average Working/Obedience Intelligence	26.0
Brightest Dogs	1.0
Excellent Working Dogs	5.0
Fair Working/Obedience Intelligence	41.0
Lowest Working/Obedience Intelligence	81.0

Once the relationship between the lower reps and percentages obeyed has been identified, I wanted to see how the average values looked in a side-by-side side comparison for each classification. As seen in figure 3, the averages also show this same kind of relationship. For

example, the dogs rated as “Excellent” had an average obeyed percent of 85 and the number of reps needed was only five. While the lowest intelligence dogs had an average percent obeyed of 30, and needed an average 81 reps.

Figure 4

Correlation:
Repetitions, Weights, Heights, and Percentages



Finally, I looked at the correlations (figure 4) between the lower/upper breed weights, the lower/ upper breed heights, the lower/upper repetitions needed, and the percent the dogs obeyed first commands. This reaffirms the relationships between the lower repetitions and the obeyed percentages. There is also a small relationship between dog weights and repetitions, but this isn’t as strong.

Conclusion. The final conclusion is that we can make recommendations on dog breeds based on their classification. Correlations were the most successful way of finding the matching breeds, which makes this the best option for gathering recommendations. By looking at similar dog breeds, we can decide which additional dog breeds might be a good match for an individual or family looking for a new dog.

Assumptions. Because different judges did the behavioral testing for different breeds, I’ve made the assumption that all dogs were assessed equally. It’s is possible that some judges may have biases for or against certain breeds; however, this wasn’t documented and I’ve assumed it hasn’t played a role in the results.

Limitations/Challenges. The greatest challenge with using these datasets were their limited sizes.. This made it difficult to split the data into meaningful test and training sets. Because of this any predictions I was able to make were meager at best. The dataset was also missing some breeds due to incomplete information.

Future Uses/Additional Applications. These recommendations could be used at shelters to assist people with finding their best dog match.

Recommendations. I think this data could be more broadly applied if it were completed with the full breed list from the AKC. Further behavioral testing would need to be done to round our the data, but once finished, this could be a useful tool.

Implementation Plan. This wasn't designed for a specific business, which means implementation plans may differ. Ideally, this would be used by a rescue to help identify the intelligence and how likely their dogs will be to obey new owners. They could also use this as a guide for knowing which animals might need more training before they go to a new home.

Ethical Assessment. The only ethics concern I have regarding this data is how the dogs might have been treated during the assessment phases. Given that the the testing was conducted by dog breed judges, I think it's probably safe to assume that these were healthy dogs that were treated well (Aerts).

Angie Oehler
DSC680
Project Milestone 3

Appendix

Aerts, C. (2014). Stanley Coren's Dog Intelligence Ranking. Retrieved from https://www.yummypets.com/mag/2014/06/24/39237/stanley-coren-dogs-intelligence-ranking

American Kennel Club. (2023). Dog Breeds. Retrieved from https://www.akc.org/dog-breeds/

Fishman, L. (2019). Dogs Intelligence and Size. Retrieved from https://www.kaggle.com/datasets/thedevastator/canine-intelligence-and-size?select=dog_intelligence.csv

VCA Animal Hospitals. (2023). Dog Breeds. Retrieved from https://vcahospitals.com/know-your-pet/dog-breeds