

Good Dogs of Cambridge

Willie Boag
MIT
wboag@mit.edu

Abstract

*The city of Cambridge is home to over 105,000 residents, yet there are only barely registered 5,000 dogs. With more than a 20-to-1 ratio of people-to-dogs, it is now more important than ever that we use big data in order to automatically identify where to find the best dogs. Because Cambridge is really h*ckin big, it can be incredibly costly to go to a neighborhood that lacks sufficient good dogs, only to realize you need to go elsewhere if you wanna be part of a high-quality belly rub. In this work, we propose a method for identifying which parts of Cambridge can maximize one's time with some big ole puppies.*

1. Background

We Rate Dogs is a Twitter account that rates pictures of dogs on a 1-10 scale. Since its inception in 2015, the account has gained over 2.2 million followers. Although most posts are of really good dogs, things were not always this way. The first dog ever rated, while still a very good pup, received a score of 8/10, as shown in Figure 1. This is not a slight against the nameless dog hero, but the low rating can certainly be understood as a reflection of the time. Nowadays, a rating like 8 is virtually unheard of. Nowadays, most ratings are either 12/10 or 13/10.

To date, @dog_rates has 6,762 posts, thus providing a rich dataset of thousands of gold truth annotations. By harnessing this data, we can model what would make my pet a really good boy. In this day and age, such an application of technology is essentially essential.

2. Experiments

We downloaded a subset of tweets from We Rate Dogs [1] in order to build a Machine Learning model. We selected a perfect number greater than 10 but less than 450, because we hypothesize that perfect numbers

Here we have a Japanese Irish Setter. Lost eye in Vietnam (?). Big fan of relaxing on stair. 8/10 would pet



Figure 1: The first dog ever rated by We Rate Dogs.

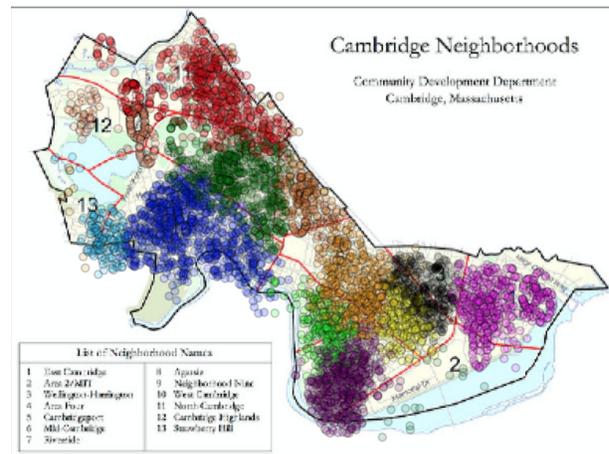


Figure 2: Every dog in Cambridge, colored by neighborhood.

are ideal for training as ML datasets. With our dozens of ratings, we extracted features from the posts to build

Neighborhood	NUM DOGS	MIN RATING	MAX RATING	AVG RATING	STD DEV
Riverside	236	11.742	14.116	12.788	0.29082
Mid-Cambridge	473	11.643	13.842	12.792	0.29723
Neighborhood Nine	649	11.529	14.305	12.800	0.33020
Wellington-Harrington	265	11.943	14.221	12.805	0.30322
Strawberry Hill	149	11.743	13.732	12.807	0.27163
Cambridge Highlands	187	11.643	13.803	12.812	0.29644
West Cambridge	852	11.743	14.305	12.819	0.31117
NA	40	12.065	13.450	12.820	0.26757
Agassiz	203	11.743	13.732	12.824	0.30788
Cambridgeport	526	11.392	14.116	12.825	0.29554
East Cambridge	508	11.682	14.249	12.828	0.29184
The Port	251	12.065	14.116	12.830	0.28150
Area 2/MIT	16	11.975	13.256	12.834	0.32300
North Cambridge	716	11.483	15.339	12.837	0.32425

Table 1: Statistics of dog ratings for each neighborhood.

a predictor that estimates 1-10 how good a given dog is. We extracted the following features:

1. length (in characters) of name.
2. All character-level unigrams, bigrams, trigrams, and quadgrams in the name.

Using scikit-learn’s Ridge [2] (i.e. L2-regularized linear regression) with $\alpha = 0.0001$, we predict a real number indicating the dog’s 1-10 rating. The code for this can be found at https://github.com/wboag/good_dogs.

We apply this predictor to a dataset of all publicly registered dogs in Cambridge, MA which is made available from the city’s Open Data initiative at <https://data.cambridgema.gov/General-Government/Dogs-of-Cambridge/sckh-3xyx>. This dataset contains the name, breed, home latitude/longitude, and neighborhood of all 5,072 registered dogs in Cambridge. Figure 2 shows the density of dogs in Cambridge by neighborhood. We can see that the three neighborhoods with the most dogs are West Cambridge (blue) with 852 dogs, North Cambridge (red) with 716 dogs, and Neighborhood Nine (green) with 649 dogs.

3. Results

We were thrilled to see how many good dogs there are in Cambridge! We observed something truly exceptional: one dog in North Cambridge named Jacob was rated a record-high 15.339/10. After careful analysis, we found the training datapoint which Jacob was most similar to: a really good boy – also named Jacob – as shown in Figure 3.



Figure 3: The training point from the WeRateDogs dataset that identified the best dog in Cambridge.

Even when we lose, we win. The lowest-rated dog in all of Cambridge had an 11.392/10, and she is from Cambridgeport.

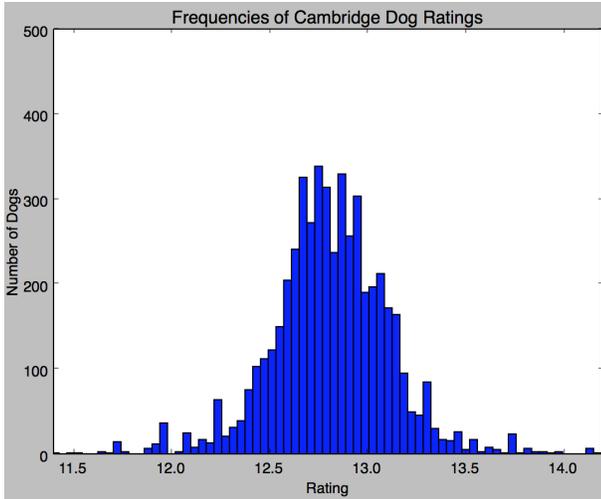


Figure 4: Histogram of ratings.

3.1. Quantitative

Cambridge’s doggos are consistently very good. The average rating is 12.817, with a standard deviation of .307. Empirically, we see that the distribution is close to Gaussian, as shown in Figure 4.

We consider the dogs in each neighborhood of Cambridge, shown in Table 1. We exclude "N/A" from our analysis because it means that the neighborhood was not tagged; for future work, we could infer the neighborhood using a 20-nearest neighbor predictor. We can see that there are so many dogs in West Cambridge, which is very exciting. On the other hand, if you would prefer a never-be-disappointed experience, every dog at The Port has a rating of at least 12.065. In contrast, for those of you that prefer consistency, Strawberry Hill has the lowest variance in their dog ratings. But for those who are feeling lucky, the authors recommend a trip to North Cambridge: not only do the dogs there have the highest average rating, but that neighborhood is also the home of Jacob, the best boy in all of Cambridge.

It is worth nothing that even though it does not have many in total, the dogs of MIT are very high quality, second only to North Cambridge. This agrees with what the authors intuitively know to be true.

3.2. Qualitative

Because 95% of the puppies fall between ratings of 12.182 and 13.451, we take an outlier-based approach to analyze the best and the not best of our furry friends. The first and fifth quintiles are shown in Figures 6 and 5, respectively.

4. Conclusions

The state of our dogs of our city is strong. Applying a probably state-of-the-art machine learning tool to rating the Dogs of Cambridge has shown that across the board, every dog is a good dog! In particular, North Cambridge seems to have an especially good group, so for your next jog, the science suggests that you should run by the parks in North Cambridge to see the finest that Cambridge has to offer.

References

[1] @dog_rates. We rate dogs.
 [2] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

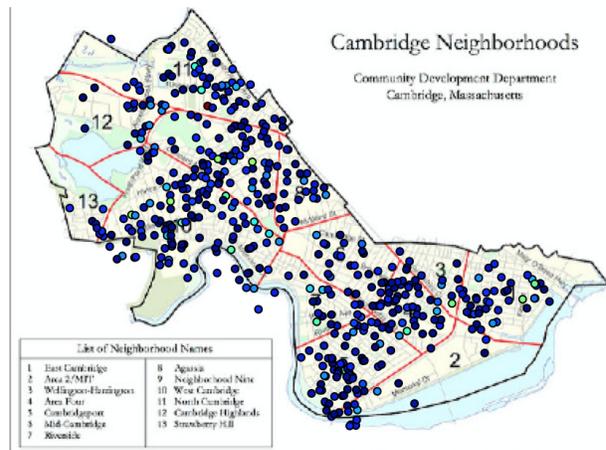


Figure 5: Top 20% of dogs in Cambridge, by rating.

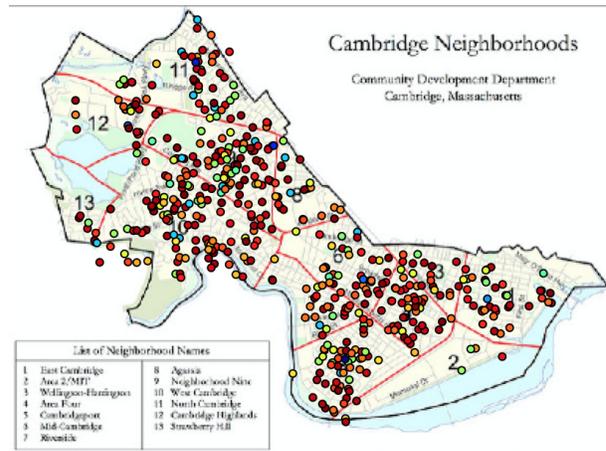


Figure 6: Bottom 20% of dogs in Cambridge, by rating.