Big Data In Animal Shelters: Using Data Science to Aid Dog Adoption

Samantha Cohen[1,2] & Peter Todd[2]

[1]: Social and Decision Analytics Division, Biocomplexity Institute, University of Virginia

[2]: Cognitive Science Program, Indiana University

Motivation

Shelters are filled with extensive data on their animals–how can we leverage this data to improve animal outcomes?

Scan me

This poster outlines work conducted at Indiana University, and published in *Behavior Research Methods* in 2019. Scan the QR code to read the paper without a subscription.

What Data Do Animal Shelters Produce?

- Visitation Data—Typical experience, peak visiting hours/days
- Shelter Management Software—Environment variables influencing adoption rates, movement throughout the shelter, photos, health
- Online Matching Services (e.g. Petfinder®)—Photos, lengths of stay, text descriptions; engagement by users, out of area interest
- Behavioral evaluations—Response of an animal to situations and stimuli, "personality" and temperament

Method

Administrative Data and Revealed Preferences

- Worked with a Midwestern animal shelter in 2017-2018.
- Used data stored in existing animal management software, including behavior, appearance, and length of stay.
- Length of stay is an indicator of *revealed* demand—dogs with more preferred traits are adopted fastest.
- Linear regression predicting length of stay based 1060 stays of 904 dogs based on traits stored in Match-Up II® .

Stated Preferences and Experimental Data

- In same animal shelter, incoming canine visitors completed a voluntary survey of *stated* preferences for an adoptable dog (n = 1229).
- Survey consisted of 13 traits, where participants could state either no preference, or up to x-1 of x trait levels.

Adopted Dog Ratings

- 145 individuals decided to adopt a dog after visiting.
- These individuals rated their chosen dogs on 45 traits on a 1 to 5 Likert scale.
- 102 of these individuals also had associated Match-Up II records.

Results

What traits in administrative data predict length of stay?

Variable	β Coefficient	*SE*	*t*	P>|t|
Constant	14.862	1.623	9.159	**>0.001**
Friendliness	-0.1659	0.072	-2.306	0.021
Playfulness	1.6813	0.331	5.077	**>0.001**
Age	0.2553	0.135	1.886	0.06
Large (Dummy)	3.5507	0.897	3.958	**>0.001**
Small (Dummy)	-2.5819	1.291	-2.0	0.046
Excitability	-0.101	0.167	-0.605	0.546
Total Commands Followed	0.0075	0.24	0.031	0.975
Fearfulness	0.1945	0.139	1.395	0.163
Dark Coloring (Dummy)	-0.6674	0.861	-0.775	0.439
Medium Coloring (Dummy)	-0.1969	0.983	-0.2	0.841
Male (Dummy)	1.0554	0.751	1.405	0.16
Purebred (Dummy)	-0.0154	0.832	-0.019	0.985
Stray (Dummy)	-3.0355	0.938	-3.235	**0.001**

Administrative data suggests most dog traits have little influence on length of stay.

What traits do visitors say they prefer?

Trait	% w/ Preference	Trait Levels		
Friendliness	88%	Not at all: 0.0%	Somewhat: 24%	Very: 83%
Playfulness	75%	Not at all: 1%	Somewhat: 69%	Very: 37%
Energy Level	74%	Low: 23%	Medium: 86%	High: 8%
Age (Years)	73%	Young (< 2): 65%	Adult (2 - 7): 45%	Senior (8+): 7%
Size	73%	Small: 34%	Medium: 59%	Large: 33%
Easily Excitable	69%	Not at all: 19%	Somewhat: 68%	Very: 17%
Intelligence	67%	Not at all: 0.4%	Somewhat: 57%	Very: 53%
Protectiveness	66%	Not at all: 16%	Somewhat: 72%	Very: 18%
Nervousness	58%	Not at all: 48%	Somewhat: 57%	Very: 1%
Previous Training	56%	No training: 8%	Some training: 91%	Extensive training: 9%
Sex	36%	Female: 56%	Male: 44%	
Purebred Status	28%	Purebred: 21%	Mixed Breed: 79%	
Color	14%	Light (White, Grey, Tan): 26%	Medium (Light Brown, Red): 46%	Dark (Black, Dark Brown): 39%

On surveys, visitors do have preferences on these traits. Preferences were not always linear nor universal.

Are administrative data good predictors of visitor perceptions?

Adopter Perceived Trait	Match-Up II Calculated Trait	Spearman's *r*	*p*	*n*
Is anxious	Fearfulness	-0.23	0.02	103
Is easily excitable	Excitability	0.32	**0.0009**	103
Is playful	Playfulness	-0.002	0.98	103
Is sociable	Friendliness	0.15	0.14	102
Is well-trained	Total Commands Followed	-0.05	0.60	103
Barks frequently	Number of subtests where 'bark' was exhibited.	0.08	0.43	103
Ignores commands	Total Commands Followed	-0.07	0.45	102

Ratings by actual adopters do not align well with those in administrative data.

Summary

Because human preferences are non-linear and idiosyncratic, typical linear regression models may mask the influence of traits on adoption. in addition, it is unclear that administrative data aligns with visitor experiences, limiting existing data's potential role in matching visitors to adoptable dogs.

UNIVERSITY of VIRGINIA
BIOCOMPLEXITY INSTITUTE

Data Science *for the* Public Good