



Tibbles & Bits

Predicting Pet Food Prices Using Text-based Product Descriptions

Team members: Simran Kathuria, Rosella Li, David Parks, Lizhou Wang

Client: Colgate-Palmolive, Dr. Iraklis (Kli) Pappas





SPECIALIZED solutions

- expert nutrition for lifestyle, breed & size
- precise formulas for common health needs
- weight-management diets



Labeling Available. Consult with Science

- Small Breed: supports immunity with antioxidants blend & small kibble
- Large Breed: supports healthy cartilage & joints with glucosamine & chondroitin
- Puppy: healthy brain & eye development with DHA from fish oil
- Adult & Adult 7+: nutrition tailored for each lifestyle



Problem Statement

Predict the sales price of new pet food products and provide related insights into the drivers of their price using the product's text-based attributes.

- If a product's price exceeds the consumer's perceived value, the product will not sell.
- If the price is below the consumer's perceived value, there is a risk of degrading brand value.

Hill's Pet Nutrition

Parent Company:	Colgate
Headquarters:	Topeka, Kansas
Annual Revenue:	\$2.2 Billion



Problem Statement

Predict the sales price of new pet food products and provide related insights into the drivers of their price using the product's text-based attributes.

- If a product's price exceeds the consumer's perceived value, the product will not sell.
- If the price is below the consumer's perceived value, there is a risk of degrading brand value.

Hill's Pet Nutrition

Parent Company: Colgate
Headquarters: Topeka, Kansas
Annual Revenue: \$2.2 Billion

$$\text{\textit{\$2.2 Billion}} \times 0.5\% = \text{\textit{\$11 Million}}$$



The Dataset

8754 observations of new pet food products introduced to the US marketplace over the past 10 years.

10 Text-based Features

- Product
- Subcategory
- Description
- Ingredients list
- ...

3 Numerical Features

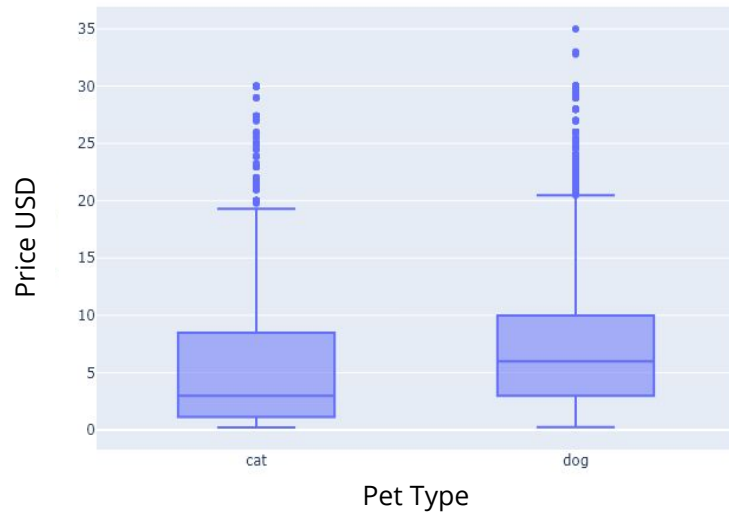
- Package size
- Unit size
- Price



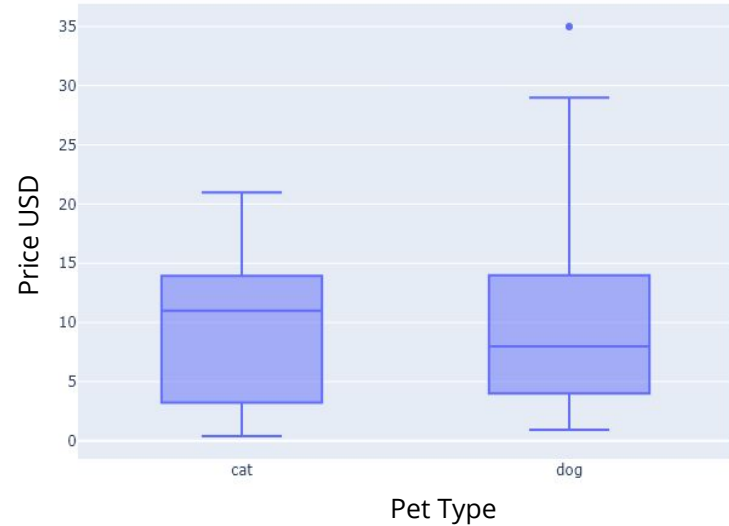
<http://allthingspups.com>

Pet Food Market Price Points

All Brands



Colgate Brands Only



Colgate maintains a higher brand value than the combined national brands.

Components

Two Separate Models

- Price prediction model
 - Take the product descriptions, numerically vectorize them, and regresses the resulting matrix against a price vector to establish the relationship between a product's description and its price.
- Grouping model
 - Group similar product descriptions to ascertain summary statistics for the group.

Literature Review

Qualitative analysis of pet owner surveys

- The same trends we see in human nutrition are starting to influence food labeling for dogs and cats. (Bohrer, 2011)
- The presence of natural ingredients was considered to be the most important indicator of pet food quality. (Vinassa, 2020)
- Most pet owners reported giving equal or more priority to buying healthy food for their pets than for themselves. (Schleicher, 2019)

Solution Methods

Possible Pricing Model Solution Methods

- Compared 19 different regression models using PyCaret
- PyCaret is an open source, low-code Python wrapper around multiple existing machine learning libraries
- Regularized linear regression performed as well as or better than more complicated regressors

The PyCaret logo, featuring the word "PYCARET" in a stylized font. "PY" is in blue, and "CARET" is in black.The spaCy logo, featuring the word "spaCy" in a blue, lowercase, sans-serif font.

Solution Methods

Possible Grouping Model Solution Methods

- K-Means clustering
- PCA
- Doc2Vec



Preferred Solutions

Predictive Pricing Model

- Term Frequency - Inverse Document Frequency (TF-IDF)
- LASSO regression

Grouping Model

- Doc2Vec
- Cosine similarity



<https://www.referee.com/pick-your-own-pregame-adventure/this-way-that-way-another-way-road-signs/>

Grouping Model Solutions

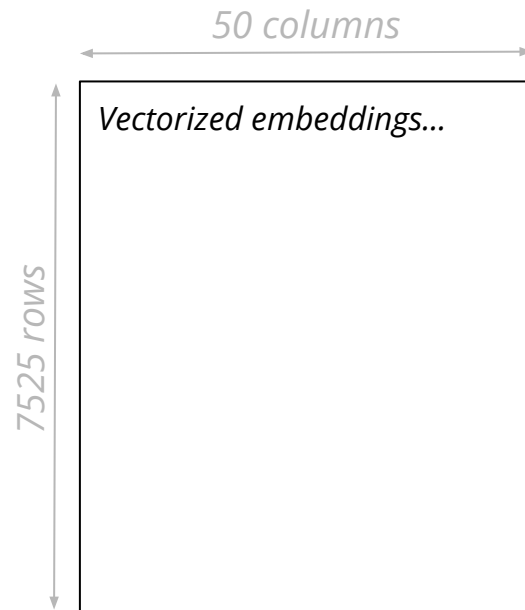
Grouping Model using DOC2VEC

Similar to Word2Vec, but Doc2Vec automatically encodes full lines of text.

Built-in cosine similarity comparison.

Model hyperparameters

- Length 50 vectors
- 40 training epochs
- Ignores words that appear fewer than 5 times in the entire corpus



Grouping Model Results for Similar Descriptions

Test description

"Tibbles & Bits brand organic dog food is naturally gluten and grain-free, and provides your pet with the best variety of raw ingredients. Follow us on Facebook to see all of our cookies, chews, and jerky."

Top match

"Robbie Dawg Organic Dog Biscuits are available in a Puppy Love Liver and Bacon variety. The veterinarian approved dog snacks contain no added salt, sugar, preservatives, artificial ingredients, wheat, corn, soy, wheat gluten or rice proteins. The organic biscuits are said to be good for dogs with allergies and sensitive digestive systems. This product is made from premium ingredients and is retailed in a 10-oz. Pack."

Top 5 Most Similar Company, Brand, and Cosine Similarity Score

Company	Brand	Similar Score
Robbie Dawg	Robbie Dawg	0.588106
Suplementos Para Mascotas Munoz Magana	Canine Chews	0.565771
Stella & Chewy's	Stella & Chewy's Carnivore Crunch	0.555846
In Clover	InClover Research	0.550339
Exclusively Pet	Exclusively Dog Cookies Best Buddy Bones	0.517176

Top 10 Most Similar Products

Summary Statistics for Price (in USD)	
Mean	9.0630
Std	6.9605
Min	1.7900
25%	2.9400
50%	7.4900
75%	14.5350
Max	20.9900

Predictive Pricing Model

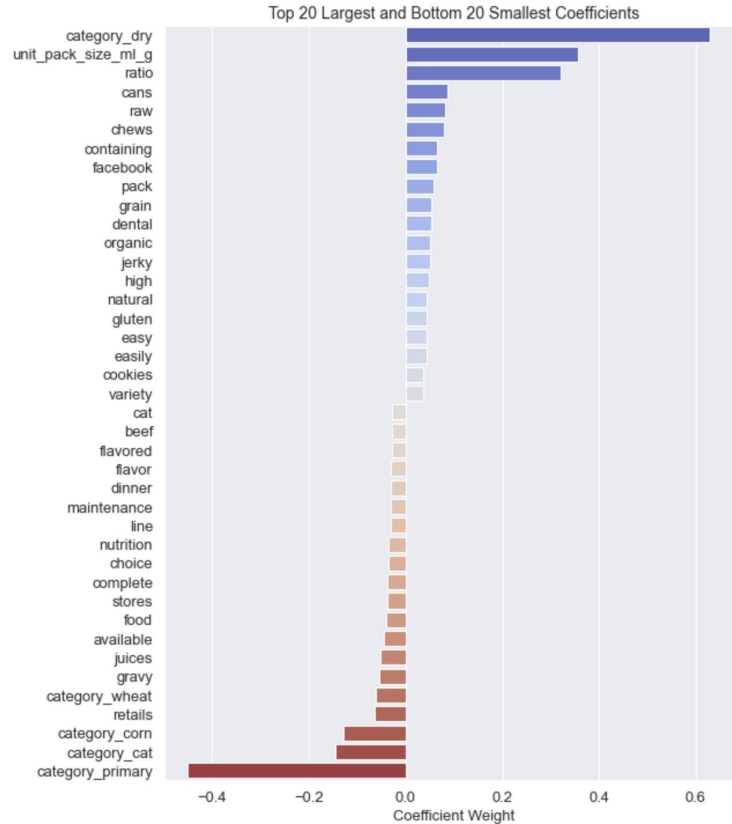
Predictive Pricing Model Results

Predictive Pricing Model Optimization Summary			
Model Type	Grid Search Hyperparameters	Optimal Hyperparameters	Root Mean Squared Error
Linear Regression	Degree: [1, ..., 20] Intercept: [True, False] Normalize: [True, False]	Degree: 1 Intercept: True Normalize: False	2.0412
Ridge Regression	Alpha: [10^{-6} , ..., 10^4] Intercept: [True, False]	Alpha: 100 Intercept: True	2.0397
LASSO Regression	Alpha: [10^{-6} , ..., 10^4] Intercept: [True, False]	Alpha: 0.006428 Intercept: True	2.0368
Elastic Net Regression	Alpha: [10^{-6} , ..., 10^4] Intercept: [True, False] L1 Ratio: [0.01, ..., 0.99]	Alpha: 0.01 Intercept: True L1 Ratio: 0.60	2.0367
Random Forest Regressor	Max depth: [10, ..., 100] Max features: [auto, sqrt] Min samples leaf: [3, ..., 50] N_estimators: [100, ..., 1000]	Max depth: 45 Max features: auto Min samples leaf: 3 N_estimators: 700	2.9028

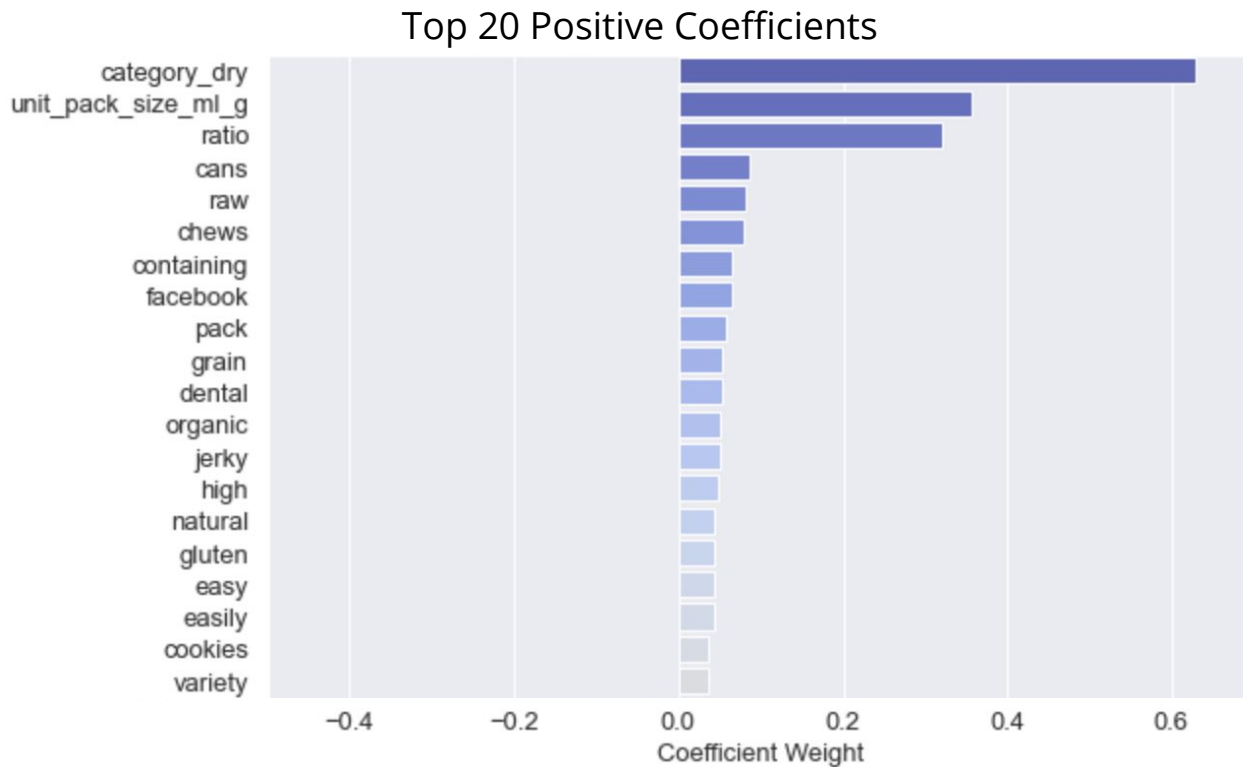
Predictive Pricing Model Results

Predictive Pricing Model Optimization Summary			
Model Type	Grid Search Hyperparameters	Optimal Hyperparameters	Root Mean Squared Error
Linear Regression	Degree: [1, ..., 20] Intercept: [True, False] Normalize: [True, False]	Degree: 1 Intercept: True Normalize: False	2.0412
Ridge Regression	Alpha: [10^{-6} , ..., 10^4] Intercept: [True, False]	Alpha: 100 Intercept: True	2.0397
LASSO Regression	Alpha: [10^{-6} , ..., 10^4] Intercept: [True, False]	Alpha: 0.006428 Intercept: True	2.0368
Elastic Net Regression	Alpha: [10^{-6} , ..., 10^4] Intercept: [True, False] L1 Ratio: [0.01, ..., 0.99]	Alpha: 0.01 Intercept: True L1 Ratio: 0.60	2.0367
Random Forest Regressor	Max depth: [10, ..., 100] Max features: [auto, sqrt] Min samples leaf: [3, ..., 50] N_estimators: [100, ..., 1000]	Max depth: 45 Max features: auto Min samples leaf: 3 N_estimators: 700	2.9028

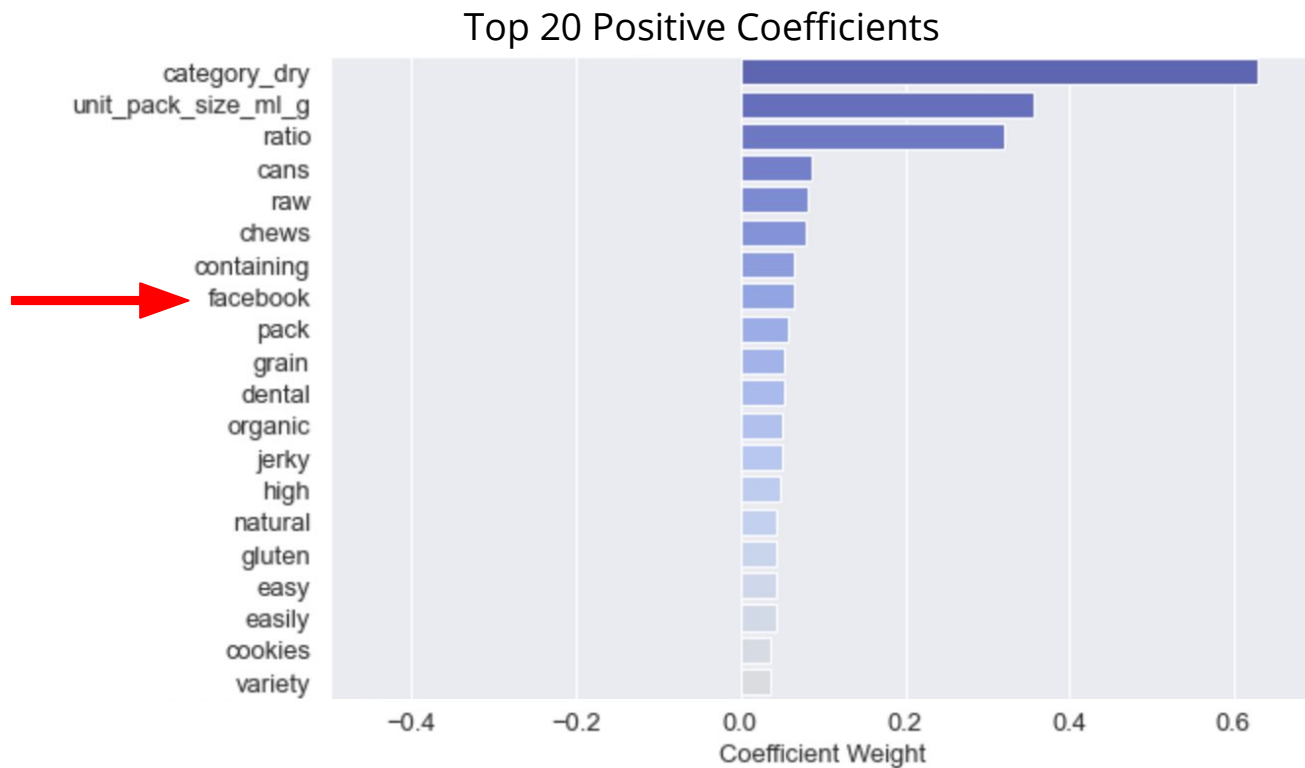
Coefficient Weights



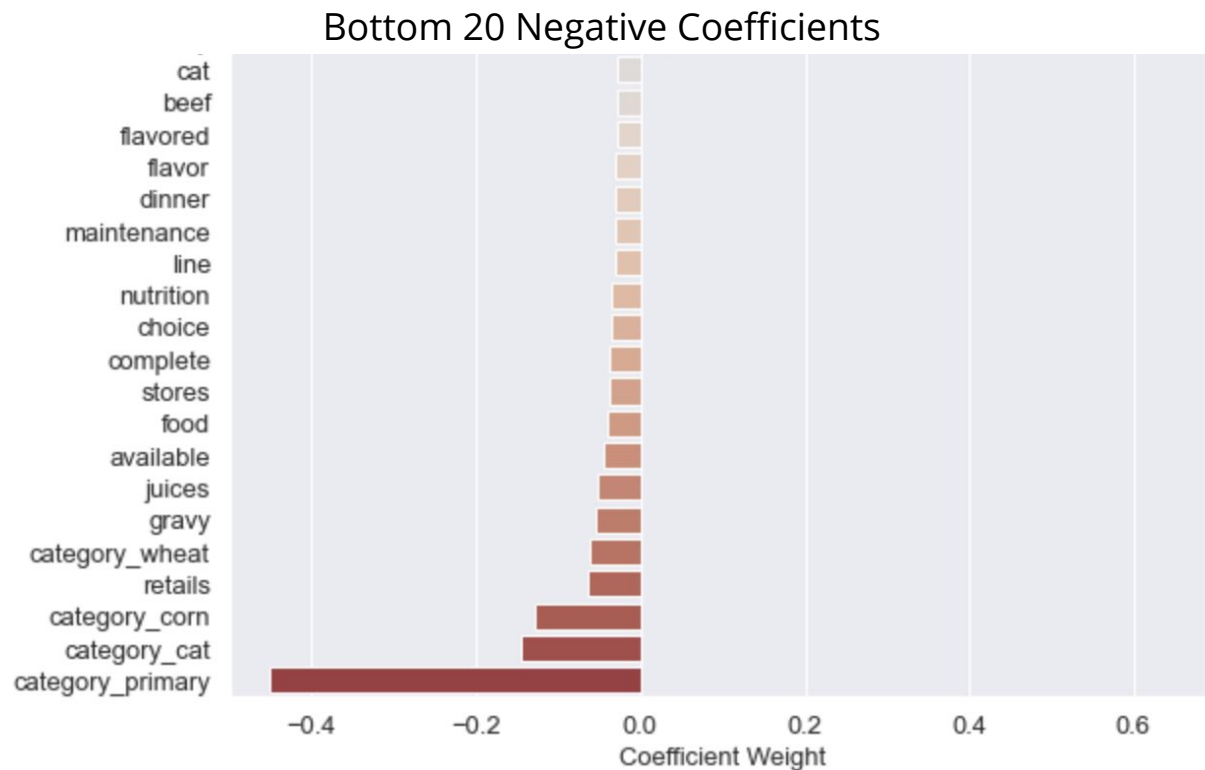
Top 20 Coefficients



Top 20 Coefficients



Bottom 20 Coefficients



Valuing Colgate's Pet Food Descriptions

The team used the model to generate predicted prices for Colgate's pet food products.

- 131 products
 - 63 over predicted prices
 - 68 under predicted prices

$$\text{sum}(\text{actual_prices} - \text{predicted_prices}) = \$110.68$$

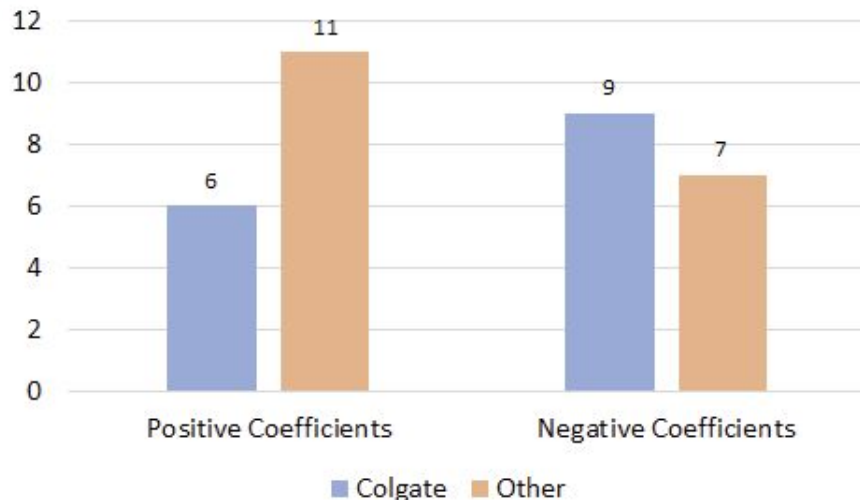
Summing the differences between the actual prices and the predicted prices produced a positive sum of \$110.68.

- In aggregate, Colgate's pet food descriptions tend to indicate a lower price relative to the actual price Colgate charges.

Why are Colgate's Pet Food Descriptions Undervalued?

1. Colgate's pet food descriptions "punch below their weight" because they include fewer trendier words and more generally descriptive words.
2. Consumers are willing to pay more for pet foods when they are recommended by a veterinarian.

*Comparison of Mean TF-IDF Scores for the Top 20 and Bottom 20 Coefficients**



* Excludes the categorical features

Limitations

Instability issues with feature matrix

- TF-IDF limitations

Assumes causal link between product description and price

- Product packaging is the consumer's first interaction with a product

Cat and dog foods only



<https://unsplash.com/photos/dQLgop4tnsc>

Lessons Learned

Data cleaning - Deeper insight

More frequent communication with client

More efficient time management

Appreciation for trying different models



<https://aflua.com/lessons-online-2014/>

Conclusions

Colgate's pet food descriptions predict a lower price relative to the actual price due to vocabulary that is less trendy and more generally descriptive.

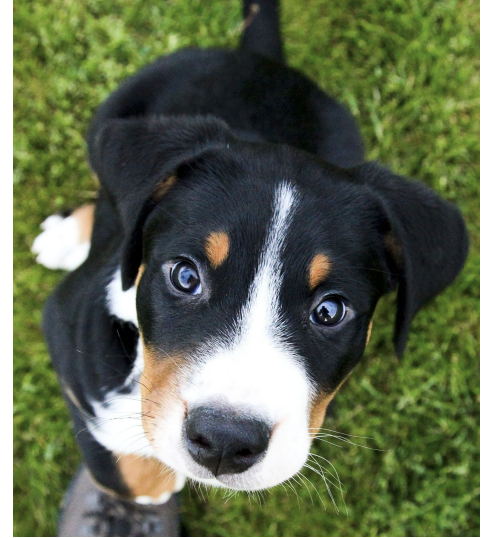
- Consider expanding vocabulary to reflect current trends.

The presence of the word “Facebook” in descriptions adds value to the overall price.

- Recommend increasing both social media engagement and the promotion of this engagement in product descriptions.

First impressions matter.

- Hill's Pet Nutrition packaging tends to be less visually appealing than some newer brands.
- Descriptions are very “sciency.”



<https://images.app.goo.gl/ZeQnpXTBY6jayCee6>