

The Quest For Five Stars

Allison Schlissel
Project Five Star
August 18, 2016



AMAZON CHALLENGES

1. Which types of food yield five star reviews?
2. How do customers describe foods in five star reviews?

amazon



DATA

- 5,000 random five-star reviews from amazon.com: Grocery & Gourmet
- MongoDB
- Pandas DataFrame



PRELIM ANALYSIS

- NLTK sentence tokenizer
- NLTK word tokenizer
- NLTK parts of speech
- Bi-Grams
- LDA

```
print_top_words(lda, tf_feature_names, n_top_words)
```

```
Topic 0:  
good really flavor best like taste ve tried  
Topic 1:
```

```
In [90]: bigram_vectorizer.get_feature_names()
```

```
Out[90]: [u'12 separate',  
          u'90 seconds']
```

K MEANS

- Dense TF/IDF Matrix
- 5 clusters
- Top 10 words in each cluster
- D3

0		1		2		3		4	
tea	0.500917	like	0.185544	food	0.285815	coffee	0.440068	great	0.080278
green	0.077957	taste	0.091269	dog	0.229588	love	0.075641	love	0.061866
like	0.058515	good	0.090046	love	0.064090	cup	0.066448	best	0.052919
good	0.053536	flavor	0.080205	quality	0.048810	good	0.061043	good	0.041646
drink	0.050219	really	0.061211	good	0.048777	great	0.058022	find	0.035732

D3? WHAT A TREAT

Let's Go!

CONCLUSIONS

- Best sellers: coffee, tea, dog food, candy
- Sweet
- Healthy
- Reasonably priced



THANK YOU

