

AMAZON FOOD REVIEWS

Aaron Osher, Amanda Short, Michael Wade



INTRO

Goal:

- To predict ratings from reviews
- To understand similar products that may be purchased or viewed
- To gain a better understanding of user behavior

Why:

- Text can be used to infer a missing rating
- An understanding of similar products aids in advertising efforts
- Using big data in a real world context



DATA

Amazon Review Dataset

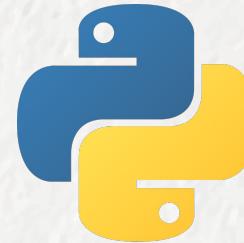
- 233 Million Amazon product reviews
- Includes product metadata
- The entire dataset is ~50 GB

We narrowed it down to a single category:

- "Grocery and gourmet food"
- 5M reviews and 200,000 products



TOOLS



GROCERY DATA

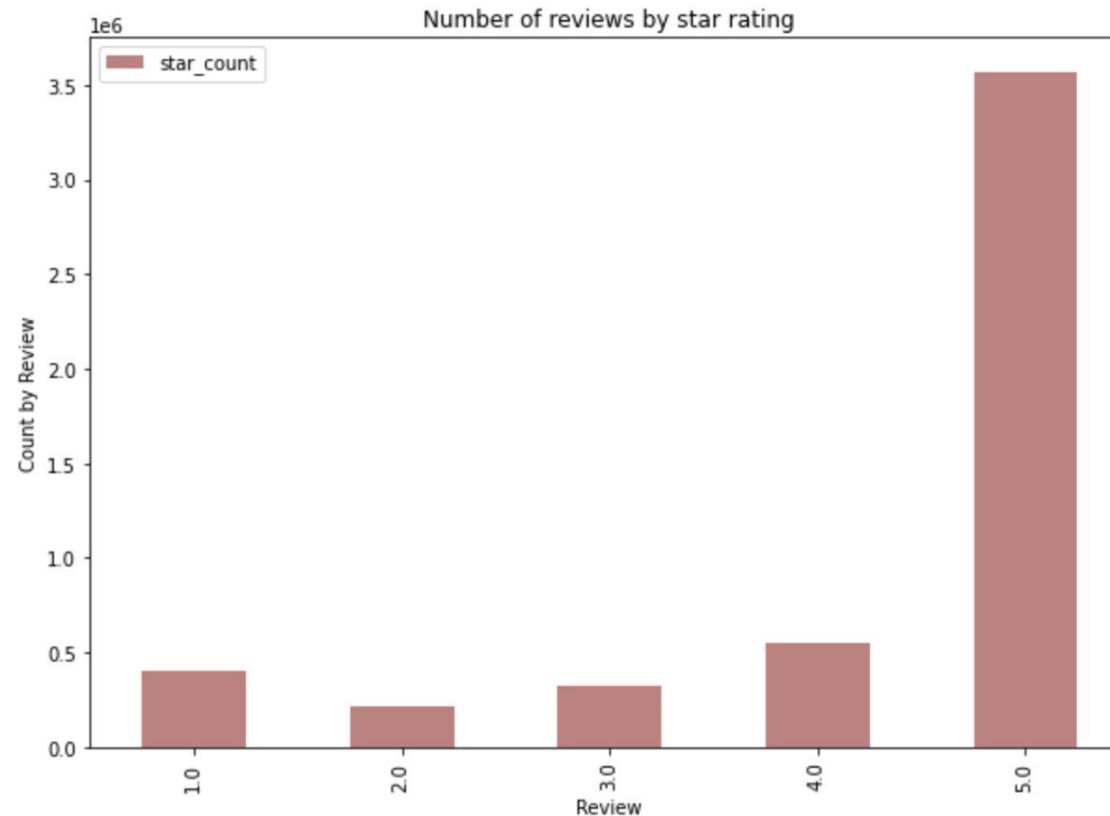


Key attributes:

- Rating
- Review text
- Summary

	asin	overall	reviewText	summary
	1888861614	5.0	Very pleased with...	Love it
	1888861614	4.0	Very nicely craft...	Nice but small
	1888861614	4.0	still very pretty... the "s" looks lik...	
	1888861614	5.0	I got this for ou... Would recommend t...	
	1888861614	4.0	It was just what ...	Topper

RATINGS



Most (3.5M) of the product reviews are 5 stars

RATING PREDICTION

Pipeline



Tokenizer (separate individual words)

Stop words & n-gram = 2

CountVectorizer (convert text to vector)

MODELS

1. Random Forest Classifier

2. **Logistic Regression**

40% accuracy in predicting ratings 1-5

84% accuracy when “binning” ratings

1-3 = “bad”, 4-5 = “good”



MODEL CHALLENGES

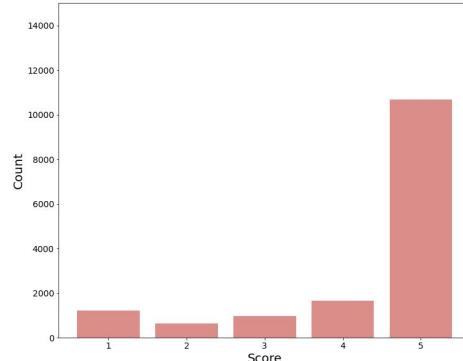
Very high class imbalance

- We downsampled to the smallest class (rating of 2)
- This led to a lower accuracy but is closer to the truth (represents all rating categories rather than just 5)

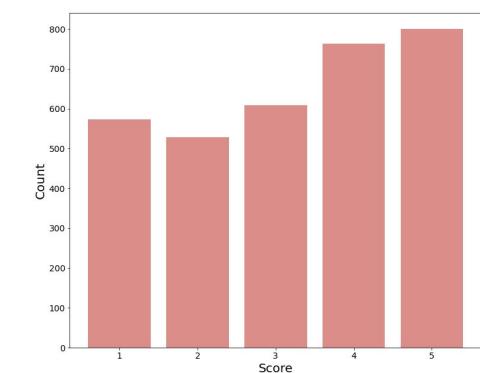
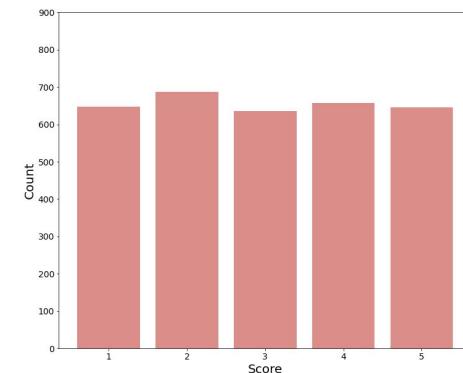
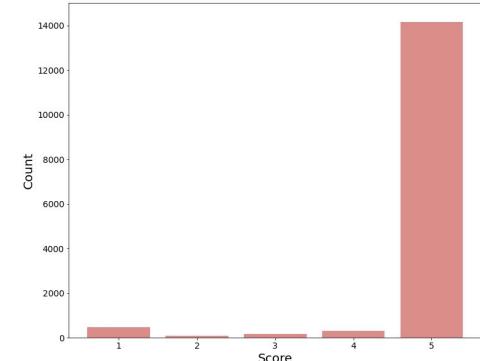
Size of the dataset

- Needed to sample for model training

Ground Truth



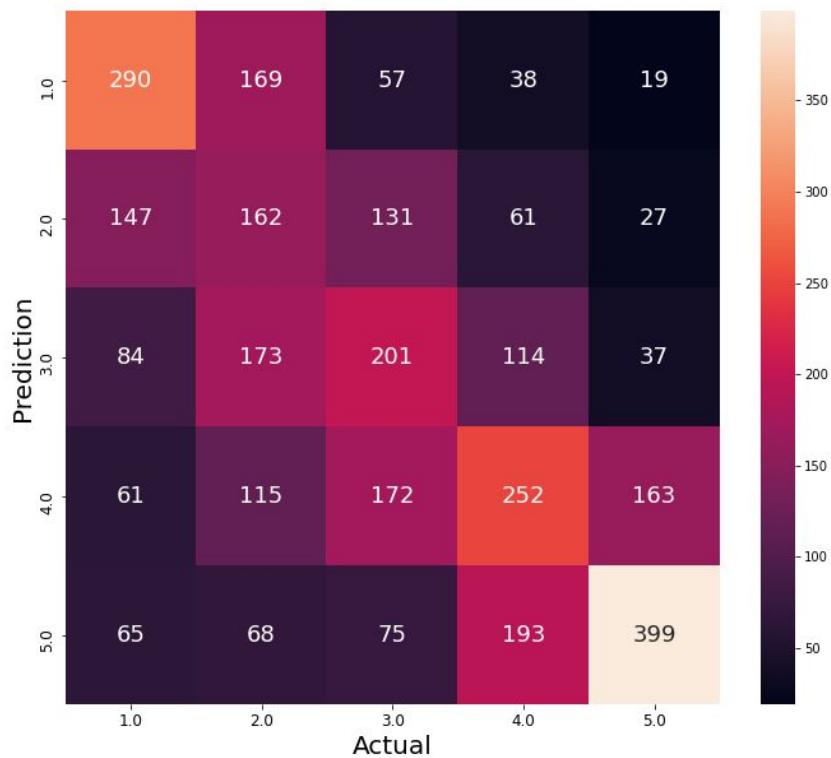
Predictions



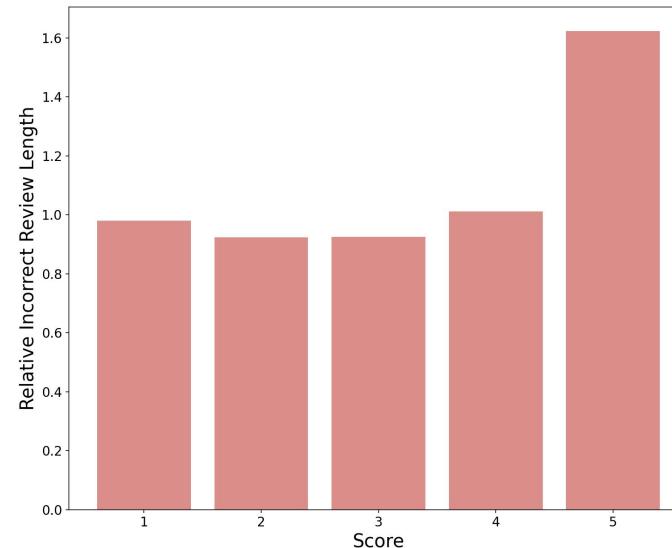
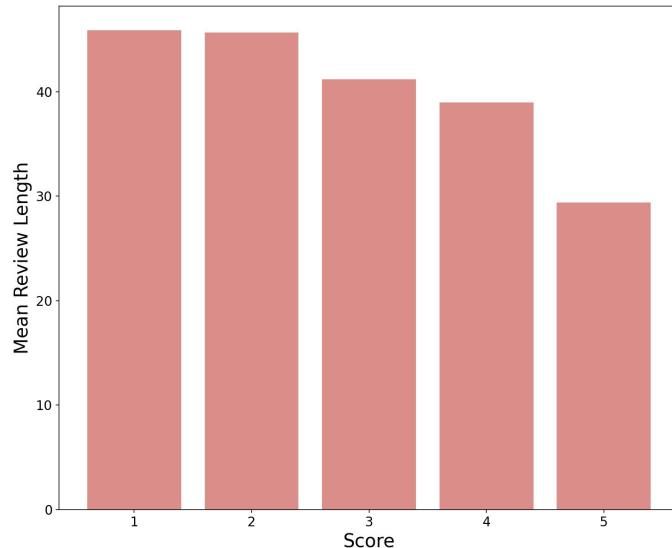
CONFUSION MATRIX

5 and 1 star ratings (the extremes) are most accurately predicted

2 and 3 stars are most difficult to predict



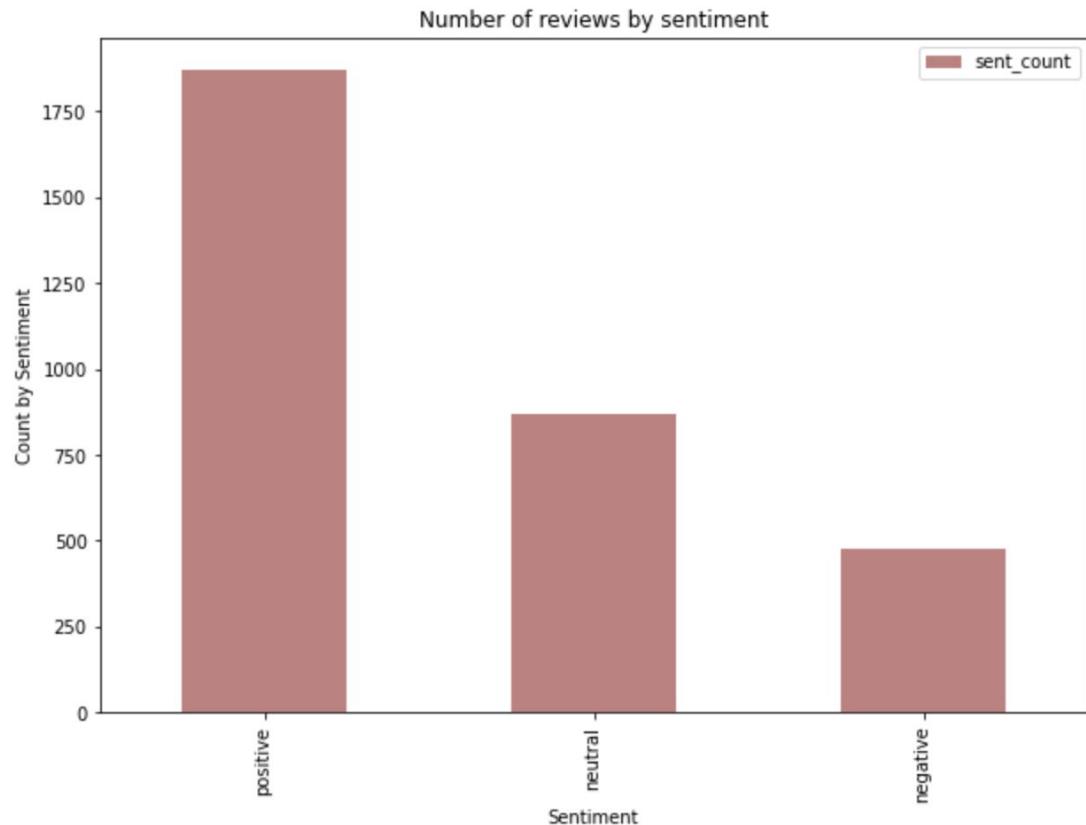
EFFECT OF REVIEW LENGTH



Longer reviews indicate worse reviews
Incorrectly predicted 5's had longer reviews than the average of the true 5 star review

SENTIMENT ANALYSIS

The model predicts that a majority of the reviews are positive, with the fewest being predicted as negative



Meta Data

Key attributes:

- Asin (product number)
- Also View
- Also Buy
- Category
- Brand

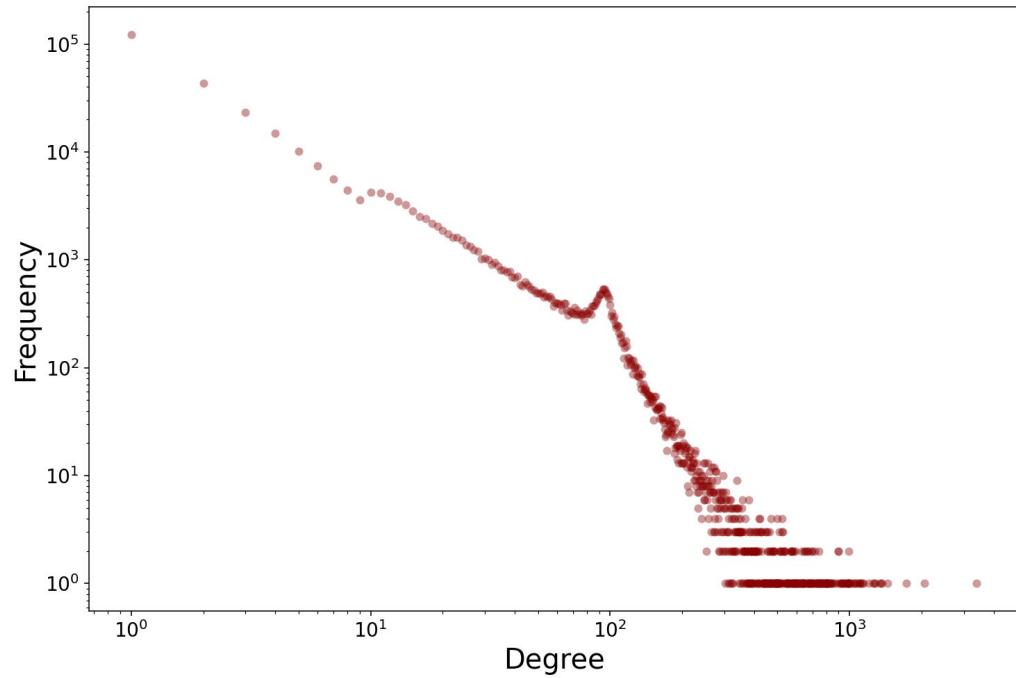
asin	also_view	also_buy	category	brand
0681727810	[B0000D9MYM, B000...	[]	[Grocery & Gourme...	Ariola Imports
0853347867		[B01898YHXX, B01B...	[Grocery & Gourme...	
1888861118	[B07DXN65TF]	[]	[Grocery & Gourme...	Unik Occasions
1888861517	[]	[]	[Grocery & Gourme...	Other
1888861614	[]	[]	[Grocery & Gourme...	Unik Occasions

NETWORK ANALYSIS

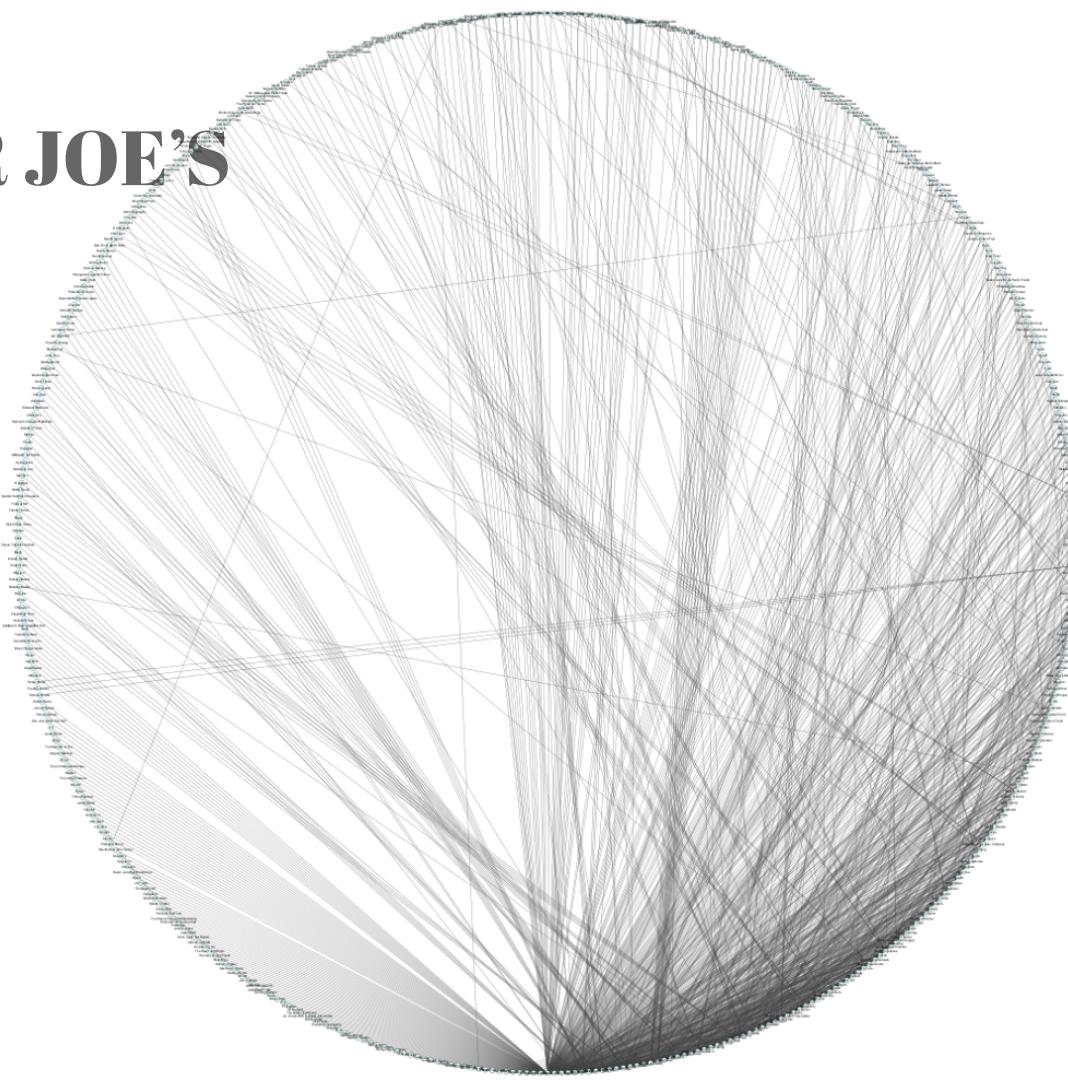
Nodes: 369,966

Edges: 1,748,020

Average degree: 4.725



TRADER JOE'S



POTENTIAL FUTURE EXPLORATION

Use summary text to predict rating rather than full text

Look at purchases across categories

Clustering



CONCLUSION

Longer text is generally associated with worse reviews

The model performs worse when dealing with longer reviews

Model performs better as a binary classifier



SOURCE

Justifying recommendations using distantly-labeled reviews and
finer-grained aspects

Jianmo Ni, Jiacheng Li, Julian McAuley

Empirical Methods in Natural Language Processing (EMNLP), 2019

