Amazon Fine Food Reviews Analysis

MSDA 683 HW 3: Exploratory Data Analysis Report

Prepared by: David Haynes

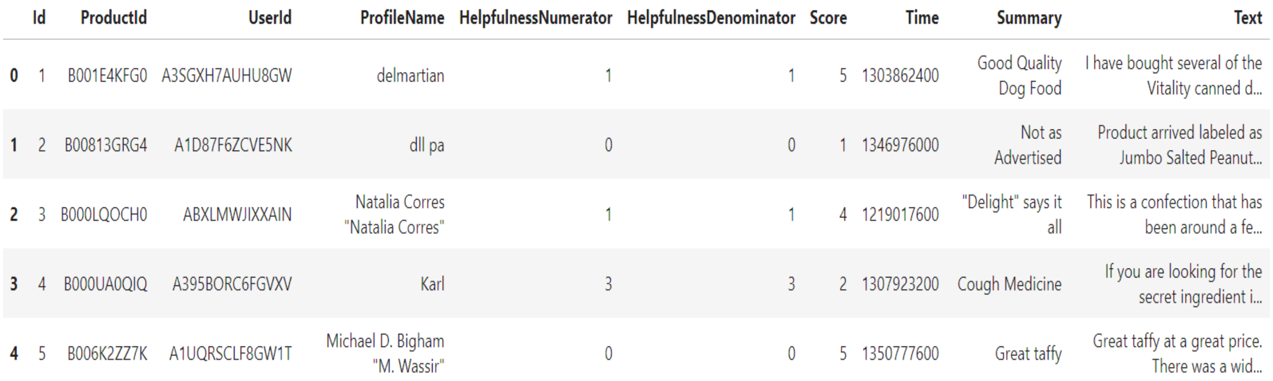
Table of Contents:

Amazon Fine Food Reviews Analysis

<table>
<tr><th colspan="2">Dataset Highlights</th></tr>
<tr><td>File Size</td><td>300.9 MB</td></tr>
<tr><td>File Format</td><td>Comma-separated value</td></tr>
<tr><td>Number of Columns</td><td>10</td></tr>
<tr><td>Column Names</td><td>Id, ProductId, UserId, ProfileName, HelpfulnessNumerator, HelpfulnessDenominator, Score, Time, Summary, Text</td></tr>
<tr><td>Total Number of Records (Rows)</td><td>568,454</td></tr>
<tr><td>Missing Attributes</td><td>ProfileName (record - 568,428), Summary (record - 568,427)</td></tr>
</table>

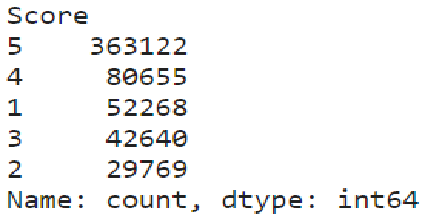
Amazon Fine Food Reviews Analysis

Dataset as a Whole:



	Id	ProductId	UserId	ProfileName	HelpfulnessNumerator	HelpfulnessDenominator	Score	Time	Summary	Text
0	1	B001E4KFG0	A3SGXH7AUHU8GW	delmartian	1	1	5	1303862400	Good Quality Dog Food	I have bought several of the Vitality canned d...
1	2	B00813GRG4	A1D87F6ZCVE5NK	dll pa	0	0	1	1346976000	Not as Advertised	Product arrived labeled as Jumbo Salted Peanut...
2	3	B000LQOCH0	ABXLMWJIXXAIN	Natalia Corres "Natalia Corres"	1	1	4	1219017600	"Delight" says it all	This is a confection that has been around a fe...
3	4	B000UA0QIQ	A395BORC6FGVXV	Karl	3	3	2	1307923200	Cough Medicine	If you are looking for the secret ingredient i...
4	5	B006K2ZZ7K	A1UQRSCLF8GW1T	Michael D. Bigham "M. Wassir"	0	0	5	1350777600	Great taffy	Great taffy at a great price. There was a wid...

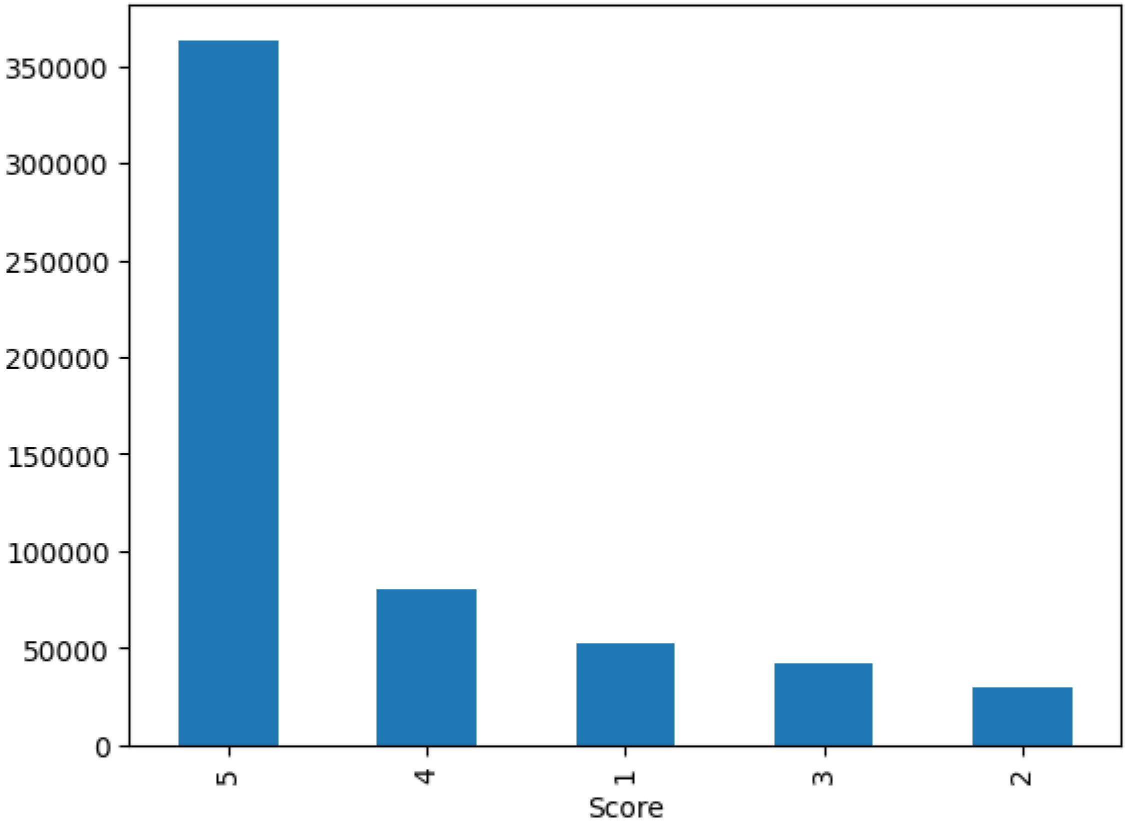
Above is a snapshot of the top 5 records in the dataset in a tabular format displaying the 10 different columns and their respective names. There are quantitative values shown under column ‘Score’ and qualitative values shown under column ‘Text’.



```
Score
5    363122
4     80655
1     52268
3     42640
2     29769
Name: count, dtype: int64
```

The figure to the left displays the breakdown of the number of reviews (records) per score. When adding up each score, the total is 568,454 which is the total number of records in the dataset. The score is in descending order in terms of the total number of reviews per score. As we can see, 5-star reviews have a vast majority of the total reviews, accounting for 64% of total reviews.

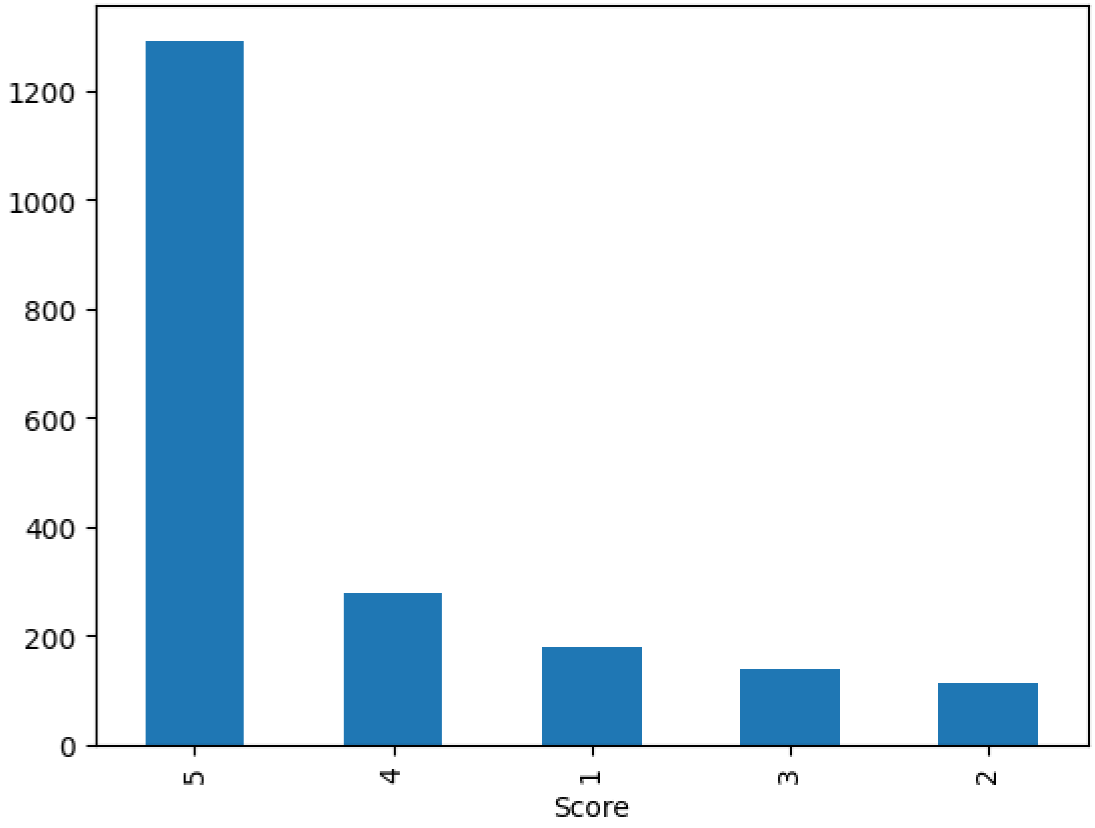
Amazon Fine Food Reviews Analysis



A bar graph visually shows the information discussed above. Five-star reviews account for a vast majority of the total reviews, followed by four-star reviews, one-star reviews, three-star reviews, and lastly 2-star reviews.

Amazon Fine Food Reviews Analysis

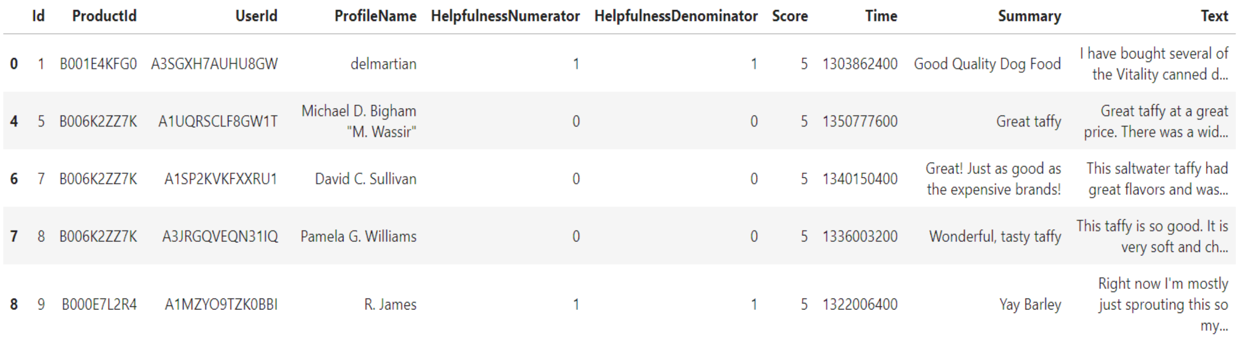
Dataset Sample:



A sample of the dataset was taken at a sample size of 2,000. The above bar chart is a representation of this sample and follows the same pattern of the dataset as a whole. Five-star reviews remain the highest number by far with more than 1,200 reviews. Four-star reviews follow with just over 200 reviews, followed by one-star reviews with just under 200 reviews. Three-star reviews come next, followed by 2-star reviews. This is a random sample and distributions are subject to change slightly.

Amazon Fine Food Reviews Analysis

Sample Five-Star Reviews:

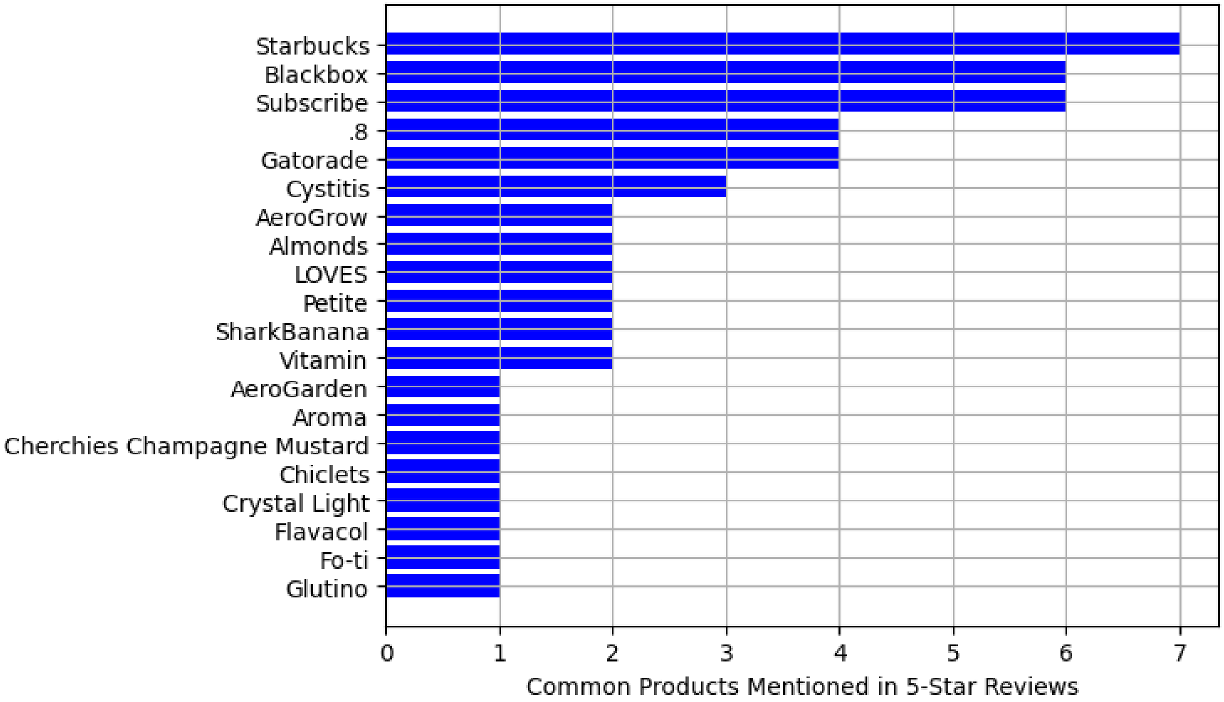


	Id	ProductId	UserId	ProfileName	HelpfulnessNumerator	HelpfulnessDenominator	Score	Time	Summary	Text
0	1	B001E4KFG0	A3SGXH7AUHU8GW	delmartian	1	1	5	1303862400	Good Quality Dog Food	I have bought several of the Vitality canned d...
4	5	B006K2ZZ7K	A1UQRSCLF8GW1T	Michael D. Bigham "M. Wassir"	0	0	5	1350777600	Great taffy	Great taffy at a great price. There was a wid...
6	7	B006K2ZZ7K	A1SP2KVKFXXRU1	David C. Sullivan	0	0	5	1340150400	Great! Just as good as the expensive brands!	This saltwater taffy had great flavors and was...
7	8	B006K2ZZ7K	A3JRGQVEQN31IQ	Pamela G. Williams	0	0	5	1336003200	Wonderful, tasty taffy	This taffy is so good. It is very soft and ch...
8	9	B000E7L2R4	A1MZYO9TZK0BBI	R. James	1	1	5	1322006400	Yay Barley	Right now I'm mostly just sprouting this so my...

great
product
like
tea
good
best
food
can
really
get
price
use
coffee
even
one
flavor
taste
little
love
something
will
also
just
much
time
amazon
healthy
many
definitely
highly
try
chocolate
eat
order
wonderful
delicious
quality
enough
make
loves
dogs
every
first
favorite
tried
bought
dog
better
buy
water
fresh
used
snack
think
easy
nice
organic
drink
perfect
bag
sweet
stuff
real
made
now
far
milk
strong
sure
right
green
brands
popcorn
around
low
recommend
cup
day
find
found
still
know
cookies
way
small
natural
free
ingredients
happy
two
store
hot
bit
years
since
treat
box
sugar
tastes
high
came

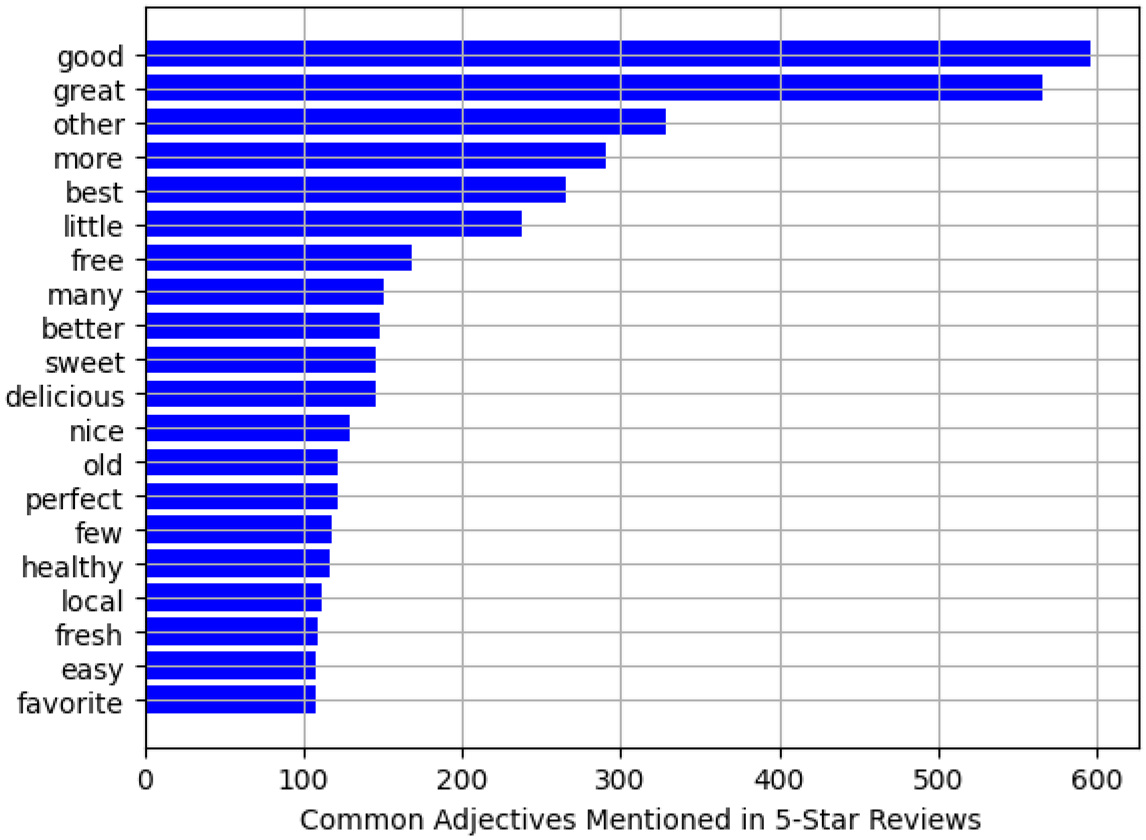
Shown above is a snapshot of the top 5 records of the sample dataset that only contain five-star scores (reviews). Pictured left is a word cloud on common words for five-star reviews pertaining to the sample data.

Amazon Fine Food Reviews Analysis



The qualitative data for the 'text' column for these five-star reviews were analyzed and graphed above via a horizontal bar chart. The chart pertains specifically to the most common products in five-star reviews. The most common product in five-star reviews is Starbucks, followed by Blackbox and Subscribe (although Subscribe may need to be filtered out in further analysis). Please note, common words that do not add value were filtered out of this analysis, as well as any other word type that does not pertain to a product.

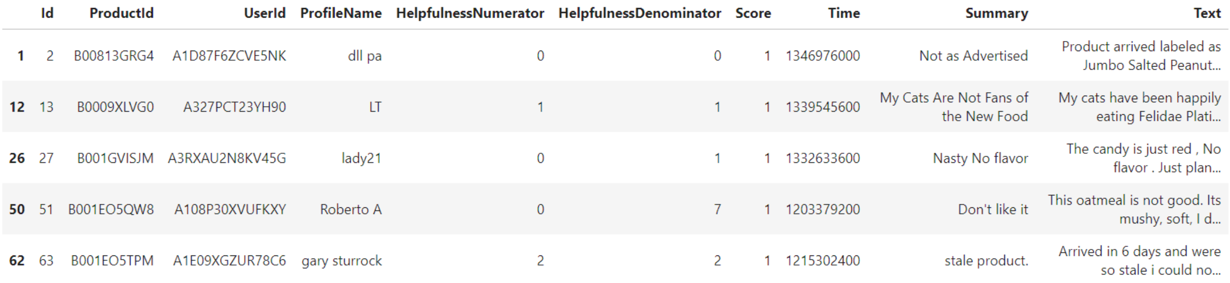
Amazon Fine Food Reviews Analysis



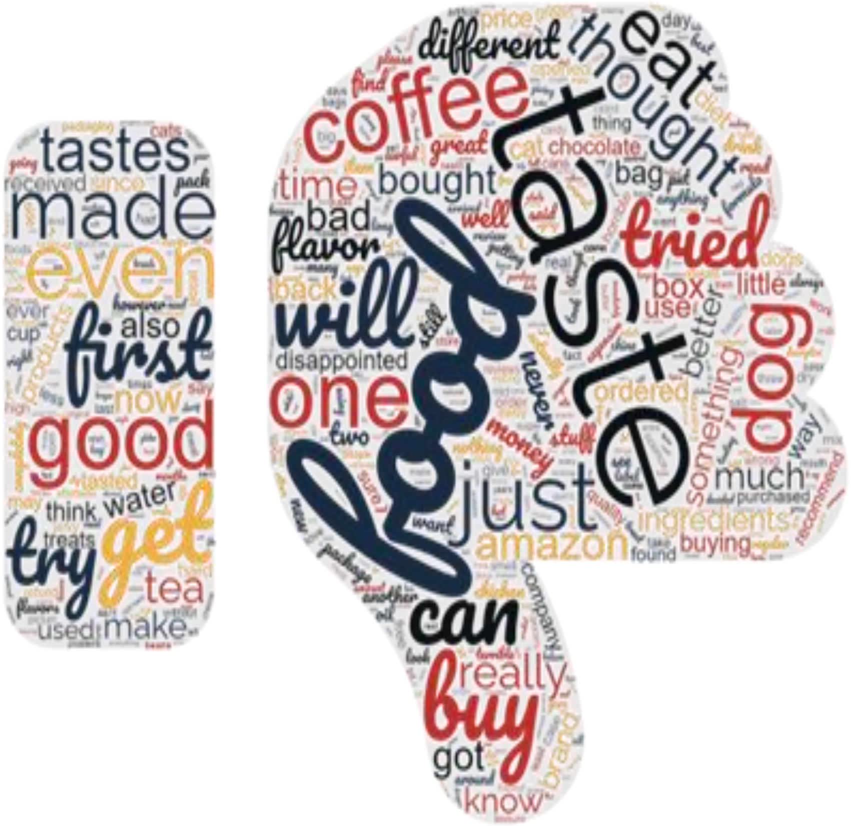
An analysis was conducted on most common adjectives in five-star reviews. Likewise with products, common words that do not add value were filtered out of this analysis, as well as any other word type that does not pertain to an adjective. The analysis shows the most common adjective in five-star reviews is good, followed by great, other, and more (other may need to be filtered out).

Amazon Fine Food Reviews Analysis

Sample One-Star Reviews:

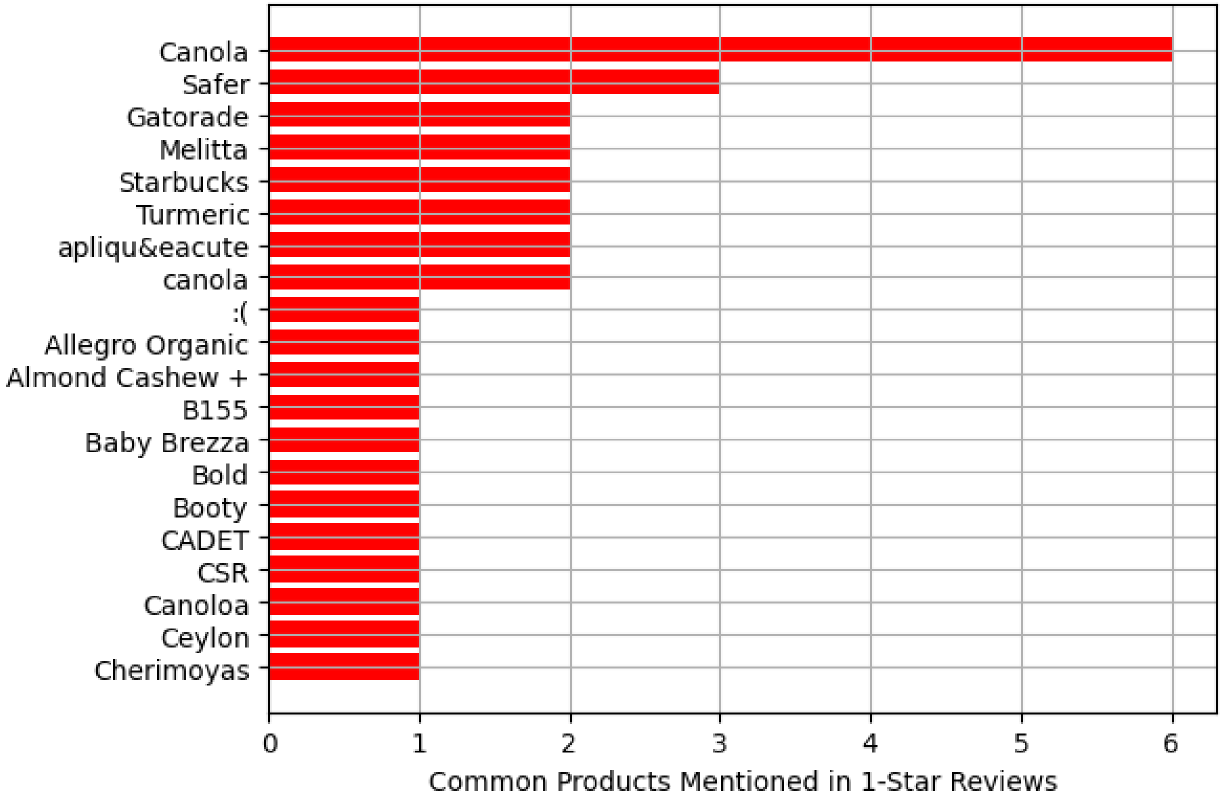


	Id	ProductId	UserId	ProfileName	HelpfulnessNumerator	HelpfulnessDenominator	Score	Time	Summary	Text
1	2	B00813GRG4	A1D87F6ZCVE5NK	dll pa	0	0	1	1346976000	Not as Advertised	Product arrived labeled as Jumbo Salted Peanut...
12	13	B0009XLVG0	A327PCT23YH90	LT	1	1	1	1339545600	My Cats Are Not Fans of the New Food	My cats have been happily eating Felidae Plati...
26	27	B001GVISJM	A3RXAU2N8KV45G	lady21	0	1	1	1332633600	Nasty No flavor	The candy is just red , No flavor . Just plan...
50	51	B001EO5QW8	A108P30XVUFKXY	Roberto A	0	7	1	1203379200	Don't like it	This oatmeal is not good. Its mushy, soft, I d...
62	63	B001EO5TPM	A1E09XGZUR78C6	gary sturrock	2	2	1	1215302400	stale product.	Arrived in 6 days and were so stale i could no...



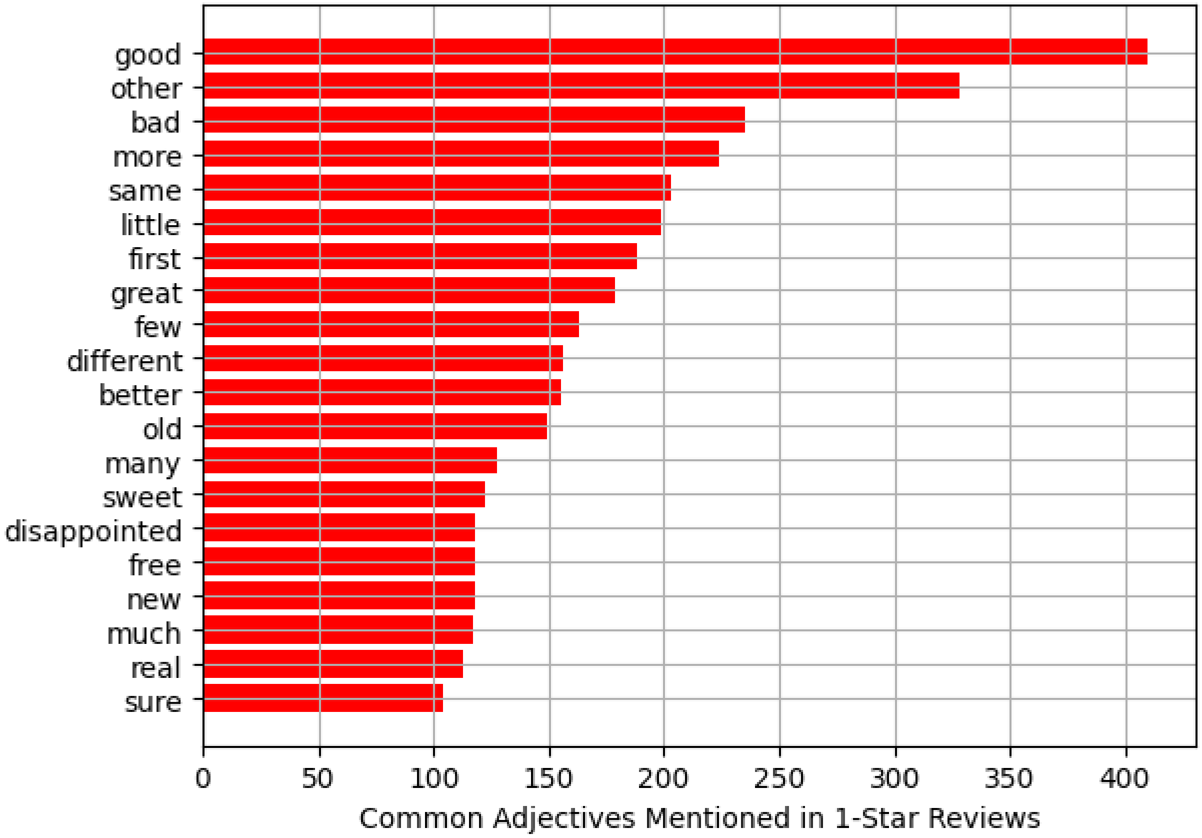
Shown above is a snapshot of the top 5 records of the sample dataset pertaining to one-star scores (reviews). Pictured left is a word cloud on common words for one-star reviews pertaining to the sample data.

Amazon Fine Food Reviews Analysis



Analysis on the most common one-star reviews in terms of products displayed above visually shows Canola, Safer, and Gatorade represent the most common products associated with one-star reviews. Starbucks is also tied for third on this list and is the leading product in five-star reviews. This insight is very interesting as is a leading contributor to request/conduct further analysis. Again, common words and non-product word types were filtered out.

Amazon Fine Food Reviews Analysis



Just as with five-star reviews, an analysis was conducted on most common adjectives in one-star reviews. Common words/non-adjective words filtered out. The analysis shows the most common adjective in one-star reviews is good, followed by other, bad, and more (other may need to be filtered out). When thinking about one-star reviews, the word good doesn't necessarily come to mind and is shockingly the most common. Additional analysis would be needed for a more depth understanding.