
NLP Project

SQL Analysis

Amazon Reviews Analysis: Synopsis Report

This project is an analysis of Amazon customer reviews, combining **Natural Language Processing (NLP) techniques and SQL queries** to extract insights from the review data. The analysis is structured into several key stages:

1. Project Setup and Data Loading

Dependency Installation: The project begins by installing the **necessary Python libraries**, primarily **nlTK for NLP tasks**.

NLTK Setup: The **NLTK downloader** is used to obtain **essential resources**, such as stop word lists.

Library Imports: The project imports libraries like **pandas for data manipulation**, **numpy for numerical operations**, and **sqlite3** to handle the **SQLite database**.

SQL Extension Loading: The **SQL extension is loaded** to enable the **execution of SQL queries** within the Jupyter notebook.

Database Connection: A connection is established with the SQLite database file (database.sqlite) containing the review data.

Initial Data Exploration: SQL queries are used to fetch and display a sample of the first 10 rows from the Reviews table, providing an initial look at the data structure and content.

2. Text Analysis

Review Fetching: SQL queries are used to retrieve the Text column, which contains the actual review text, from the Reviews table.

Text Conversion and Tokenization: The review texts are converted to lowercase and combined into a single string. The text is then **tokenized (split into individual words) using regular expressions**.

Word Frequency Calculation: The frequency of each word is calculated using the **Counter Algorithm function**.

DataFrame Creation: The word counts are converted into a Pandas DataFrame, with columns labeled "word" and "count," and are sorted by frequency in descending order. The top 100 words are displayed.

3. Basic Data Analysis using SQL

Total Review Count: A SQL query is executed to count the total number of reviews in the Reviews table.

Score Distribution: SQL queries are used to determine the **frequency of each review score**, providing insights into the overall sentiment distribution.

Oldest and Latest Reviews: SQL queries are employed to retrieve the **oldest and latest reviews** based on the Time column, allowing for a **chronological examination** of the reviews.

4. Review Text Exploration

The project delves into examining the content of individual reviews, displaying snippets of both old and new reviews.

This provides context for the **word frequency analysis** and helps in understanding the nature of the reviews.

5. Advanced Text Analysis

Missing Review Check:

The project checks for any missing reviews by counting the number of entries where the Text column is null or empty.

Review Length Analysis:

Reviews are categorized based on their length (Very Short, Short, Medium, Long), and the count of reviews in each category is calculated.

Peak Review Activity:

The project identifies peak review activity by grouping reviews by the Time column and counting the number of reviews for each time period.

Data Transformation:

The Time column (Unix timestamp) is converted into a more readable datetime format.

Stop Word Removal and Text Cleaning:

Punctuation is removed from the **review text**, and **stop words** (common words like "the," "is," "and") are eliminated to focus on **meaningful words**.

Text Preprocessing Function:

A function **text_process** is defined to perform the punctuation removal and stop word removal.

Applying the Function:

The text_process function is applied to the Text column of the reviews DataFrame.

Data Exploration:

The project explores the structure and information of the reviews DataFrame using reviews.info().

6. Sentiment Analysis and Visualization

Sentiment Analysis:

The TextBlob library is used to perform sentiment analysis on the review text.

Visualization:

The project includes various visualizations, such as histograms and bar plots, to illustrate data characteristics like review length distribution and transaction details.

SQL Analysis Summary

The "Amazon Reviews Analysis" project provides a comprehensive exploration of Amazon customer review data. It combines SQL for data retrieval and manipulation with Python and NLP techniques for text analysis. The project covers essential steps from data loading and cleaning to text processing, frequency analysis, and sentiment analysis, providing a solid foundation for **understanding customer opinions and trends within the review dataset**.

Downloading The Dependencies

```
In [3]: pip install nltk
```

```
Requirement already satisfied: nltk in c:\users\collins pc\anaconda3\envs\collonel\lib\site-packages (3.8.1)
Requirement already satisfied: click in c:\users\collins pc\anaconda3\envs\collonel\lib\site-packages (from nltk) (8.1.7)
Requirement already satisfied: joblib in c:\users\collins pc\anaconda3\envs\collonel\lib\site-packages (from nltk) (1.4.0)
Requirement already satisfied: regex<=2021.8.3 in c:\users\collins pc\anaconda3\envs\collonel\lib\site-packages (from nltk) (2024.4.16)
Requirement already satisfied: tqdm in c:\users\collins pc\anaconda3\envs\collonel\lib\site-packages (from nltk) (4.66.2)
Requirement already satisfied: colorama in c:\users\collins pc\anaconda3\envs\collonel\lib\site-packages (from click->nltk) (0.4.6)
Note: you may need to restart the kernel to use updated packages.
```

```
In [2]: import nltk
```

```
In [7]: nltk.download_shell()
```

```
NLTK Downloader
```

```
-----
d) Download  l) List    u) Update  c) Config  h) Help    q) Quit
-----
```

```
Download which package (l=list; x=cancel)?
```

```
Downloading package stopwords to C:\Users\Collins
PC\AppData\Roaming\nltk_data...
Package stopwords is already up-to-date!
```

```
-----
d) Download  l) List    u) Update  c) Config  h) Help    q) Quit
-----
```

```
In [3]: import pandas as pd
import numpy as np
```

```
In [4]: # Load and activate the SQL extension to execute SQL in a Jupyter notebook.
%load_ext sql
```

```
In [5]: #Load the database file
# Make sure the file is saved in the same folder as this notebook.
%sql sqlite:///database.sqlite
```

```
In [20]: %%sql
SELECT
    *
FROM
    Reviews
LIMIT 10;

* sqlite:///database.sqlite
Done.
```

Out[20]:

	Id	ProductId	UserId	ProfileName	HelpfulnessNumerator	HelpfulnessDenominator	Score	Time	Summ
--	----	-----------	--------	-------------	----------------------	------------------------	-------	------	------

1	B001E4KFG0	A3SGXH7AUHU8GW	delmartian		1	1	5	1303862400	Gc Que Dog Fe
---	------------	----------------	------------	--	---	---	---	------------	---------------------

2	B00813GRG4	A1D87F6ZCVE5NK	dll pa		0	0	1	1346976000	Not Advertis
---	------------	----------------	--------	--	---	---	---	------------	-----------------

3	B000LQOCH0	ABXLMWJIXXAIN	Natalia Corres "Natalia Corres"		1	1	4	1219017600	"Delig says ii
---	------------	---------------	------------------------------------	--	---	---	---	------------	-------------------

4	B000UA0QIQ	A395BORC6FGVXV	Karl	3	3	2	1307923200	Cot Medic
---	------------	----------------	------	---	---	---	------------	-----------

5	B006K2ZZ7K	A1UQRSCLF8GW1T	Michael D. Bigham "M. Wassir"	0	0	5	1350777600	Great t
---	------------	----------------	-------------------------------	---	---	---	------------	---------

6	B006K2ZZ7K	ADT0SRK1MGOEU	Twoapennything	0	0	4	1342051200	Nice T
---	------------	---------------	----------------	---	---	---	------------	--------

7	B006K2ZZ7K	A1SP2KVKFXXRU1	David C. Sullivan	0	0	5	1340150400	Gre Just good expens bran
---	------------	----------------	----------------------	---	---	---	------------	---

8	B006K2ZZ7K	A3JRGQVEQN31IQ	Pamela G. Williams	0	0	5	1336003200	Wonder tasty t
---	------------	----------------	-----------------------	---	---	---	------------	-------------------

9	B000E7L2R4	A1MZY09TZK0BBI	R. James	1	1	5	1322006400	Yay Bar
---	------------	----------------	----------	---	---	---	------------	---------

10	B00171APVA	A21BT40VZCCYT4	Carol A. Reed	0	0	5	1351209600	Heal Dog Fo
----	------------	----------------	---------------	---	---	---	------------	----------------

```
In [21]: import sqlite3
import pandas as pd
import re
from collections import Counter

# Connect to SQLite database
conn = sqlite3.connect("database.sqlite")
cursor = conn.cursor()

# Fetch reviews
query = "SELECT Text FROM Reviews;"
cursor.execute(query)
reviews = cursor.fetchall()

# Convert to a single text list
all_text = " ".join([row[0].lower() for row in reviews])

# Tokenization: Split words using regex
words = re.findall(r'\b\w+\b', all_text)

# Count word frequency
word_counts = Counter(words)

# Convert to DataFrame
df = pd.DataFrame(word_counts.items(), columns=["word", "count"]).sort_values(by="count", ascending=False).head

print(df)
```

```
# Close connection
cursor.close()
conn.close()
```

```
      word    count
5      the  1872601
0       i  1691844
11     and  1289768
23      a  1235964
28     it  1073835
..     ...     ...
264   find   73520
185   well   73408
408   make   72007
497  after   71514
30  better   70767
```

[100 rows x 2 columns]

In [22]: df

Out[22]:

	word	count
5	the	1872601
0	i	1691844
11	and	1289768
23	a	1235964
28	it	1073835
...
264	find	73520
185	well	73408
408	make	72007
497	after	71514
30	better	70767

100 rows × 2 columns

In [23]:

```
%%sql
SELECT COUNT(*) AS total_reviews FROM Reviews;
```

* sqlite:///database.sqlite
Done.

Out[23]:

total_reviews
568454

In [24]:

```
%%sql
--Most Frequent Scores in the reviews table
SELECT Score, COUNT(*) AS count
FROM Reviews
GROUP BY Score
ORDER BY count DESC
LIMIT 200;
```

* sqlite:///database.sqlite
Done.

Out[24]:

Score	count
5	363122
4	80655
1	52268
3	42640
2	29769

In [25]:

```
%%sql
--check the oldest reviews
SELECT Text FROM Reviews
ORDER BY Time DESC
LIMIT 50;
```

* sqlite:///database.sqlite
Done.

This is a very healthy dog food. Good for their digestion. Also good for small puppies. My dog eats her required amount at every feeding.

My partner is very happy with the tea, and is feeling much better since starting to drink it.
She has been drinking it both hot (normal) and iced (chilled) and likes the refreshing nature of it.

Purchased this product at a local store in NY and my kids and i love it. Its a quick easy meal. You can put in a toaster oven on toast for 6 min. and its ready to eat.
strongly recommend

These lollipops are are well done, look exactly like the picture. Not a single one was broken! They packaged each lollipop so nicely, very impressed!

I have ordered these raisins multiple times. They are always great and arrive timely. I can't go back to store bought chocolate covered raisins now! Love this product.

I do not like sour taste and this has a sour kind of taste which i don't like. The smell isn't that great either

I just called Bob's Red Mill customer service (just do a G search for the company, their customer service number is under "contact us" on their site), and I was annoyed to learn that NONE OF BOB'S RED MILL BAKING SODA CONTAINS ALUMINUM. I've been paying \$10-15 (with shipping) for this magical 'aluminum free' baking soda only to learn that the Bob's Red Mill 'premium baking soda' at my local grocery is the same exact thing. Call for yourself if you don't believe me. Ugh.

Rodeo Drive is my absolute favorite and I'm ready to order more! That's if I can find it.
I don't know why they are discontinuing it.
It arrived very fast.

Rodeo Drive is my absolute favorite and I'm ready to order more! That's if I can find it.
I don't know why they are discontinuing it.
It arrived very fast.

Hi,
I am very disappointed with the past shipment I received of the ONE coconut water. 3 of the boxes were leaking and the coconut water was spoiled.

Thanks.

Laks

This is one of the very finest maple syrups I have ordered and enjoyed. I bought the Fancy Grade, for the light amber color. It has a very great taste. I will reorder.

I love this product and order it frequently. Usually it arrives in a timely manner and in good condition. When I opened this last shipment I found some hemp seed scattered throughout the box. I carefully removed the first of two pouches and found a 1" rip in the middle of the front of the bag. Most of the product was still in the pouch and appeared fresh and undamaged so I transferred it to a container. The other pouch was intact. I hope this is a one-time occurrence, and I will be ordering this again when I need more. More care needs to be taken to assure that the pouches are handled properly and that they are all in good condition when packed.

Rooibos Natural Red tea is something of a personal taste! It has an interesting flavor as all things that are good for you do! I prefer to add a little cinnamon to mine. It is known to be used for people who get GOUT and since I do , I have had no incidents since drinking this tea! If it works, use it! It is a good tea and caffeine free. Some prefer it with sugar!Give it a try too for it's calming effect!

This is the "all gone" treat after dinner. It's the only treat that the dogs work for; and I did run the chance of losing a hand. They know a new command now: "be gentle" when taking liver treats.

Love this faucet. My husband had installed the same one in our old house so when our current faucet was leaking I told him not to fix it and we would buy the same one. It was easy enough for him to install but he did need my assistance a few times to help hold some things in place. Looks great and works great.

Great coffee at a good price. I'm a subscription buyer and I buy this month after month. What more can I say?

Great coffee and so easy to brew. This coffee has great aroma and is good to the last drop. I actually like all the brands. This is the way coffee should taste!!

This is my everyday coffee choice...a good all around crowd pleaser. Green mountain Sumatra would be my back-up-for-a-change-of-pace second choice...nice to have both on hand!

Just made my first pot of this wonderful coffee. It was simply delish! Subtle hint of cinnamon. It wasn't overpowering, which I was worried about. I've had 2 cups already. I'll be making this lots! It's difficult to find flavored types of coffee that are decaf. Thumbs up! :-)

We drink a lot of tea... from all over the world! By far this is the worst tasting tea we have had... I purchased this based on reviews and because this is organic. When compared to st dalfour organic tea... this tastes like Lipton caffeine free tea from Costco! I HAVE always gone with reviews before buying anything online, but this particular tea is no good, At ALL. And Costco returns is better than amazon for so much of something so bad!

I got the teriyaki flavor and, while the flavor is good and the texture is also good once you get it into your mouth, it can be a mess getting it there. As other reviewers have noted, this is not like your typical dry jerky. It seems more like something that would be created by the company that makes SlimJims or something - it's moist and chewy. The individual packets don't tear easily, either, despite being notched at the end. You know how sometimes plastic can tear very cleanly, but sometimes it turns into a ragged edge and becomes a struggle? That's what happens with these 8 times out of 10. So if you're out and about (not near a sink) and you try to open it and the tear doesn't work perfectly, you potentially have liquid dripping out onto you, certainly on your fingers as you're trying to get the jerky out of the package and into your mouth. And the liquid dries sticky.

Anyway, as I said, the flavor is good and I've now eaten most of mine, but I have taken to opening the package with scissors and eating them only at home so I can wash my hands afterward.

Nespresso makes GREAT coffee and GREAT machines. I switched over to a Nespresso machine 7 years ago and have never looked back. I save a small fortune every year by making my lattes at home.

That being said, the Nespresso capsule offers posted here are from a third party who is putting a large additional margin on their price.

You can order the same products online from Nespresso for approx \$0.55 each (half the price here).

each (than the price here).

We love this drink mix it taste delicious and impossible yo find at the market in town. Everyone else loves it too! To solve the Delia now we just order a case!

Individually packaged assorted rice crackers are really good. I could without all the packaging. Really great price for a high quality mix of rice crackers with some wasabi peas included. I mixed with salted peanuts for a delicious snack.

We originally purchased these chews from our veterinarian. Decided to order them here from Amazon because of the reduced cost and lack of shipping temperature restrictions on the product. We get complimented on the whiteness of our dogs' "smiles" all the time. Our vet is happy and so am I because I don't have to brush their teeth a lot! You definitely want to wash your hands after handling the chews because they have a peculiar odor. Now if someone can figure out how to ship heart worm prevention without it getting too hot, I'd be ecstatic!

The coffee "Blend" is really lean on Kona and really fat on fill. I could not get a full bodied taste no matter how I ground it. If you want real Kona taste, spend the bucks and buy real Kona!!!

My cats love these. They are very hard to find anywhere but here. They used to be available in the pet stores but no longer. Must refrigerate after opening. They are a little pricy but a healthy treat for your cat.

I got the teriyaki flavor and, while the flavor is good and the texture is also good once you get it into your mouth, it can be a mess getting it there. As other reviewers have noted, this is not like your typical dry jerky. It seems more like something that would be created by the company that makes SlimJims or something - it's moist and chewy. The individual packets don't tear easily, either, despite being notched at the end. You know how sometimes plastic can tear very cleanly, but sometimes it turns into a ragged edge and becomes a struggle? That's what happens with these 8 times out of 10. So if you're out and about (not near a sink) and you try to open it and the tear doesn't work perfectly, you potentially have liquid dripping out onto you, certainly on your fingers as you're trying to get the jerky out of the package and into your mouth. And the liquid dries sticky.

Anyway, as I said, the flavor is good and I've now eaten most of mine, but I have taken to opening the package with scissors and eating them only at home so I can wash my hands afterward.

I really enjoyed these and was surprised at how much they remind me of a "normal" chocolate-oatmeal-raisin cookie. But I don't really feel they should be called nutrition bars, as while they have 28% RDA of dietary fiber per bar, they have only 4g of protein, 4% RDA of calcium and 8% RDA of iron, and that's it; no other vitamins/minerals. When I hear "nutrition bar" I'm expecting a bit more than that. Even so, they are tasty and filling and only 200 calories per bar so I'll buy some from time to time for those "rushed mornings" when a decent breakfast is impossible.

Excellent smooth taste. Any one who loves chocolate will enjoy this drink. Almost has good as International house's dark mayan chocolate coffee was.

Yum, Yum, Yum, Yum... (Munch, munch). Yum! Love it and it tastes wonderful. Works perfect in a whirly-popcorn maker. They best compliment is when the kids ask, can we make another batch, and they like it better than the movie. Our friends have tried it and have asked for what we use as our popcorn recipe. How cool is that!?

Have used this for years. Gives long lasting energy boost with no sudden let down. Good taste. 4 oz ea. so can put them in my carry-on when traveling.

Svenhard's varity pack pastries are great. Individually wrapped to stay fresh and very tasty. Box of 36 of different varities. Freezes well. I eat one every day that we are at home!

We recently returned from a wonderful three week excursion to Ireland.
We tasted many Irish teas and our favorite was Barry's Gold Blend.
I was pleased to find it readily available on Amazon and at a reasonable price
as we have already used up the supply we brought home.

I wanted a food for a a dog with skin problems. His skin greatly improved with the switch, though he still itches some. He loves the food. No recalls, American made with American ingredients was a big plus (no Chinese poisoned food). I took off a point for availability, which increases the price. Fortunately he is a small dog and the food lasts a while. He has done well so now feeling safe to expand the diet to other flavones in this line for variety.

Hubby and I (both 63) have been drinking about 4 oz a night for a few months now. We both found that we are falling into a deep sleep faster and for a longer time (eliminating the 2 or 4 am wake up). My sister and her hubby also tried it and like it very much for the same reason.Indian Summer 100% Juice, Montmorency Cherry, 46-Ounce Containers (Pack of 8)

These are small and very salty. The taste is good, but very strong, so it's a good thing the package contains a small amount. It only takes a few little crisps to cure my salty/crunchy craving. I can snack on one package for an entire day. Of course these would not be a good snack if you're very hungry, because there isn't enough there to fill you up. For less than \$1 per pack it's an okay deal.

does anyone know if the flowers come with personal message? It doesn't show where i can write a message to go with the flowers.

These are small and very salty. The taste is good, but very strong, so it's a good thing the package contains a small amount. It only takes a few little crisps to cure my salty/crunchy craving. I can snack on one package for an entire day. Of course these would not be a good snack if you're very hungry, because there isn't enough there to fill you up. For less than \$1 per pack it's an okay deal.

This stuff tasted like (insert favorite negative adjective that Amazon won't print). I like most cereals but I couldn't put this stuff down, tasted like cardboard something with a half stale texture. Hate to throw 3/4 box of cereal away, we'll see if the birds will eat it - I have my doubts.

Great for HS lunch, kid enjoy as a snack also, will buy again. Salted chips are good too, tried them too.

Nifty hot chocolate discs added to your warm milk or milk alternative is simply yum! Made in Ireland so you know the chocolate is high quality.

Ingredients for those that desire it (I think it should be a requirement on any listing):
Milk chocolate 29% (sugar, whole milk powder, cocoa butter, cocoa mass, emulsifier:soya lecithin, flavouring: natural
vanilla), Cocoa mass 25%, Invert sugar syrup, Sugar syrup, sugar, water, preservative:sorbic acid), Sweetened condensed
milk (milk, sugar). Milk chocolate contains 32% cocoa, solids minimum, 20% milk solids minimum.

Shelf life is about 18 months.

This is a favorite of mine for using over ice. Even bought it to give out as Christmas gifts last year.

I love this stuff! It's a God sent Remedy for hangovers!.... I understand the health concerns many people have and I can also see that this can be habit forming. I make sure I only take one every so often and not everyday.

Good price, flavor, fast delivery And good presentation overall. Very efficient when playing golf, it really gives extra Energy and it helps to achieve a better game

Just tried the orange and iced coffee this morning and really liked them both! I am going to place my order today!

I love popcorn, and this product is good except for the fact that it does not and will not give you the taste that you are looking for that you get from movie theaters. Even with the matching seasoning salt, it doesn't do the justice. But its ok,just not buttery at all.

First off, read the ingredients, no crazy words I can't pronounce, which means it's all natural! I got the peanut butter treats for my two children: husky/shepherd and a lab/shepherd mix, and they can't get enough of them. Plus, on Amazon, this is such a great value for 16oz. Beats petsmart for sure! I'll be looking for more Zuke's products now that I've discovered them. Thanks Zuke's!

I love the assortment. Different countries of origins. It is fun to taste the differences. So good. My favorite is the Peruvian.

So yummy... Drinking it Black coffee or w cream this coffee is delish One of my favs... I would recommend this coffee to everyone :)

```
In [26]: %%sql
--check the latest reviews
SELECT Text FROM Reviews
ORDER BY Time ASC
LIMIT 50;
```

```
* sqlite:///database.sqlite
Done.
```

Out[26]:

Text

this witty little book makes my son laugh at loud. i recite it in the car as we're driving along and he always can sing the refrain. he's learned about whales, India, drooping roses: i love all the new words this book introduces and the silliness of it all. this is a classic book i am willing to bet my son will STILL be able to recite from memory when he is in college

I can remember seeing the show when it aired on television years ago, when I was a child. My sister later bought me the LP (which I have to this day, I'm thirty something).I used this series of books & songs when I did my student teaching for preschoolers & turned the whole school on to it. I am now purchasing it on CD, along with the books for my children 5 & 2. The tradition lives on!

Beetlejuice is a well written movie everything about it is excellent! From the acting to the special effects you will be delighted you chose to view this movie.

A twist of rumplestiskin captured on film, starring michael keaton and geena davis in their prime. Tim Burton's masterpiece, rumbles with absurdity, and is wonderfully paced to the point where there is not a dull moment.

A twist of rumplestiskin captured on film, starring michael keaton and geena davis in their prime. Tim Burton's masterpiece, rumbles with absurdity, and is wonderfully paced to the point where there is not a dull moment.

A twist of rumplestiskin captured on film, starring michael keaton and geena davis in their prime. Tim Burton's masterpiece, rumbles with absurdity, and is wonderfully paced to the point where there is not a dull moment.

Beetlejuice is an excellent and funny movie. Keaton is hilarious as the wacky beetlejuice. The great special effects help the film. I think this is one of the best movies ever made and I'm sure you'll agree. For a good time, watch beetlejuice!

THIS IS ONE MOVIE THAT SHOULD BE IN YOUR MOVIE COLLECTION. IT IS FILLED WITH COMEDY, ACTION, AND WHATEVER ELSE YOU WANT TO CALL IT.

THIS IS ONE MOVIE THAT SHOULD BE IN YOUR MOVIE COLLECTION. IT IS FILLED WITH COMEDY, ACTION, AND WHATEVER ELSE YOU WANT TO CALL IT.

THIS IS ONE MOVIE THAT SHOULD BE IN YOUR MOVIE COLLECTION. IT IS FILLED WITH COMEDY, ACTION, AND WHATEVER ELSE YOU WANT TO CALL IT.

I, myself always enjoyed this movie, it's very funny and entertaining, so I didn't hesitate to pick up the clamshell edition. I guess it was a marketing plan to make the movie more for families or something, but they have eliminated all strong profanity and elements that are usually edited in the television version. YOU HAVE BEEN WARNED. If you want the uncut version, AVOID THE CLAMSHELL EDITION!

I, myself always enjoyed this movie, it's very funny and entertaining, so I didn't hesitate to pick up the clamshell edition. I guess it was a marketing plan to make the movie more for families or something, but they have eliminated all strong profanity and elements that are usually edited in the television version. YOU HAVE BEEN WARNED. If you want the uncut version, AVOID THE CLAMSHELL EDITION!

I, myself always enjoyed this movie, it's very funny and entertaining, so I didn't hesitate to pick up the clamshell edition. I guess it was a marketing plan to make the movie more for families or something, but they have eliminated all strong profanity and elements that are usually edited in the television version. YOU HAVE BEEN WARNED. If you want the uncut version, AVOID THE CLAMSHELL EDITION!

I bought a few of these after my apartment was infested with fruit flies. After only a few hours, the trap had "attracted" many flies and within a few days they were practically gone. This may not be a long term solution, but if flies are driving you crazy, consider buying this. One caution- the surface is very sticky, so try to avoid touching it.

What happens when you say his name three times? Michael Keaten stars in this comedy about two couples that live in an old two story house. While coming back from a supply store, the couple suddenly get caught inside of a "broken-up" bridge and then just before they start to tumble down into the lake, a board catches them. But just when they've got their hopes up, and small dog steps on the board and the car starts to slide off the bridge and into the lake waters. A few minutes later...<p>They find themselves back into their home, they find that somehow somehow light the fireplace, as if done by magic. From then on, they find a weird-looking dead guy known as Bettlejuice. The only way they can get him for help is to call him by his name three times and he will appear at their survice. But they soon wish that they have never called his name, because Bettlejuice was once a troublemaker but he is the only one who can save them, on the account that they said his name three times. They can't leave their houses or else they will find theirselves in another world with giant sandworms. This is a stellar comedy that you should see! Michael Keaton is awesome as he plays the leading role of Bettlejuice.

What happens when you say his name three times? Michael Keaten stars in this comedy about two couples that live in an old two story house. While coming back from a supply store, the couple suddenly get caught inside of a "broken-up" bridge and then just before they start to tumble down into the lake, a board catches them. But just when they've got their hopes up, and small dog steps on the board and the car starts to slide off the bridge and into the lake waters. A few minutes later...<p>They find themselves back into their home, they find that somehow somehow light the fireplace, as if done by magic. From then on, they find a weird-looking dead guy known as Beetlejuice. The only way they can get him for help is to call him by his name three times and he will appear at their service. But they soon wish that they have never called his name, because Beetlejuice was once a troublemaker but he is the only one who can save them, on the account that they said his name three times. They can't leave their houses or else they will find themselves in another world with giant sandworms. This is a stellar comedy that you should see! Michael Keaton is awesome as he plays the leading role of Beetlejuice.

What happens when you say his name three times? Michael Keaten stars in this comedy about two couples that live in an old two story house. While coming back from a supply store, the couple suddenly get caught inside of a "broken-up" bridge and then just before they start to tumble down into the lake, a board catches them. But just when they've got their hopes up, and small dog steps on the board and the car starts to slide off the bridge and into the lake waters. A few minutes later...<p>They find themselves back into their home, they find that somehow somehow light the fireplace, as if done by magic. From then on, they find a weird-looking dead guy known as Beetlejuice. The only way they can get him for help is to call him by his name three times and he will appear at their service. But they soon wish that they have never called his name, because Beetlejuice was once a troublemaker but he is the only one who can save them, on the account that they said his name three times. They can't leave their houses or else they will find themselves in another world with giant sandworms. This is a stellar comedy that you should see! Michael Keaton is awesome as he plays the leading role of Beetlejuice.

I'm getting crazy.I'm looking for Beetlejuice french version video.<p>Is it really impossible today not to find the French VHS version of this film ?<p>Could U please tell me something about it ? Tks

I'm getting crazy.I'm looking for Beetlejuice french version video.<p>Is it really impossible today not to find the French VHS version of this film ?<p>Could U please tell me something about it ? Tks

I'm getting crazy.<p>Is it really impossible today not to find the French VHS version of this film ?<p>Could U please tell me something about it ? Tks

This was a really good idea and the final product is outstanding. I use the decals on my car window and everybody asks where i bought the decals i made. Two thumbs up!

I just received my shipment and could hardly wait to try this product. We love "slickers" which is what we call them, instead of stickers because they can be removed so easily. My daughter designed signs to be printed in reverse to use on her car windows. They printed beautifully (we have 'The Print Shop' program). I am going to have a lot of fun with this product because there are windows everywhere and other surfaces like tv screens and computer monitors.

I have just recently purchased the Woodstream Corp 0610 gopher trap, and within 10 minutes of laying the traps i was catching gophers. This product is the best i have ever used. The 0610 gopher trap is easy to set and works with great success. Also, remember to use a wire attached to the trap and tie it to a steak to prevent gophers from dragging the trap further into their hole once they are caught. I hope you find this product as easy to use as i did, Good luck.

This are so much easier to use than the Wilson paste colors. Colors are vibrant, and do not taint the frosting like some colors can. These are simple to use, and do not make a mess. My only complaint is that I did not find these years ago. This is a must have if you decorate often!

These are easy to use, they do not make a mess, and offer vibrant colors. They do not taint what you are decorating as some colors can. I would highly recommend these to anyone who likes to decorate.

This is such a great film, I don't even know how to sum it up. First of all, it is completely original and it is unlike any film I have ever seen before. Second of all, it's a great comedy with kind of a spooky, weird feel to it, which is something all of Tim Burton's films have. The look of the film is probably what I like the best. Art Director Bo Welch and Tim Burton show us a world unlike anything seen in a movie. This is a great film, and I would recommend it to anyone looking for an enjoyable, entertaining film that is original and inventive.

This is such a great film, I don't even know how to sum it up. First of all, it is completely original and it is unlike any film I have ever seen before. Second of all, it's a great comedy with kind of a spooky, weird feel to it, which is something all of Tim Burton's films have. The look of the film is probably what I like the best. Art Director Bo Welch and Tim Burton show us a world unlike anything seen in a movie. This is a great film, and I would recommend it to anyone looking for an enjoyable, entertaining film that is original and inventive.

This is such a great film, I don't even know how to sum it up. First of all, it is completely original and it is unlike any film I have ever seen before. Second of all, it's a great comedy with kind of a spooky, weird feel to it, which is something all of Tim Burton's films have. The look of the film is probably what I like the best. Art Director Bo Welch and Tim Burton show us a world unlike anything seen in a movie. This is a great film, and I would recommend it to anyone looking for an enjoyable, entertaining film that is original and inventive.

Beetlejuice is an awe-inspiring wonderfully amusing comedic romp that explores the incredible possibilities of the after-life and the boundaries of the absurd. Telling the tale of a recently dead married couple that suddenly gets led into the chaotic world of the supernatural. Adam and Barbara Maitland (Alec Baldwin and Geena Davis) discover many of the conflicting rather human imperfections that haunt the lives of the living also plague the afterlife as well.<p>Unlike some film projects that seemed to be blindly assigned to dispassionate filmmakers for commercial reasons, Beetlejuice's plot and bizarre subject matter remarkably complemented Burton's unusually macabre artistic sensibilities extraordinarily well. Created by the unbelievably brilliant guidance of imaginative film director, Tim Burton (Pee Wee's Big Adventure, Batman, Ed Wood, Sleepy Hollow). The film's uniquely creative landscape was a culmination of essentially all of the abundant irony and outlandish yet human behaviors that have graced the body of Burton's work.<p>Augmented by deviously energetic performances from Glenn Shadix, Jeffrey Jones, Winona Ryder, Catherine O' Hara, Geena Davis, and Alec Baldwin, the film bustles with uninhibited brilliant hilarity that persistently pushes the film to a level of almost affable euphoria. Paired with its ingenious screenplay and a tour de force performance by Michael Keaton as Beetlejuice. This film transforms itself into an exuberant jovial exercise that is extremely satisfying and philosophically perceptive on many levels. Though comedy is usually considered to be unable or undeserving of deep critical analysis, Beetlejuice is an undeniably inspired concept that has been flawlessly transferred to film. One of the most outstanding comedies of the 1980's. <p>As for the film's DVD package, the DVD includes the theatrical trailer, an isolated Danny Elfman musical track, and the choice to watch the film in anamorphic widescreen or Pan and Scan. Hopefully Beetlejuice will eventually be graced with a special edition. We can always hope.

Beetlejuice is an awe-inspiring wonderfully amusing comedic romp that explores the incredible possibilities of the after-life and the boundaries of the absurd. Telling the tale of a recently dead married couple that suddenly gets led into the chaotic world of the supernatural. Adam and Barbara Maitland (Alec Baldwin and Geena Davis) discover many of the conflicting rather human imperfections that haunt the lives of the living also plague the afterlife as well.<p>Unlike some film projects that seemed to be blindly assigned to dispassionate filmmakers for commercial reasons, Beetlejuice's plot and bizarre subject matter remarkably complemented Burton's unusually macabre artistic sensibilities extraordinarily well. Created by the unbelievably brilliant guidance of imaginative film director, Tim Burton (Pee Wee's Big Adventure, Batman, Ed Wood, Sleepy Hollow). The film's uniquely creative landscape was a culmination of essentially all of the abundant irony and outlandish yet human behaviors that have graced the body of Burton's work.<p>Augmented by deviously energetic performances from Glenn Shadix, Jeffrey Jones, Winona Ryder, Catherine O' Hara, Geena Davis, and Alec Baldwin, the film bustles with uninhibited brilliant hilarity that persistently pushes the film to a level of almost affable euphoria. Paired with its ingenious screenplay and a tour de force performance by Michael Keaton as Beetlejuice. This film transforms itself into an exuberant jovial exercise that is extremely satisfying and philosophically perceptive on many levels. Though comedy is usually considered to be unable or undeserving of deep critical analysis, Beetlejuice is an undeniably inspired concept that has been flawlessly transferred to film. One of the most outstanding comedies of the 1980's. <p>As for the film's DVD package, the DVD includes the theatrical trailer, an isolated Danny Elfman musical track, and the choice to watch the film in anamorphic widescreen or Pan and Scan. Hopefully Beetlejuice will eventually be graced with a special edition. We can always hope.

Beetlejuice is an awe-inspiring wonderfully amusing comedic romp that explores the incredible possibilities of the after-life and the boundaries of the absurd. Telling the tale of a recently dead married couple that suddenly gets led into the chaotic world of the supernatural. Adam and Barbara

Maitland (Alec Baldwin and Geena Davis) discover many of the conflicting rather human imperfections that haunt the lives of the living also plague the afterlife as well.<p>Unlike some film projects that seemed to be blindly assigned to dispassionate filmmakers for commercial reasons, Beetlejuice's plot and bizarre subject matter remarkably complemented Burton's unusually macabre artistic sensibilities extraordinarily well. Created by the unbelievably brilliant guidance of imaginative film director, Tim Burton (Pee Wee's Big Adventure, Batman, Ed Wood, Sleepy Hollow). The film's uniquely creative landscape was a culmination of essentially all of the abundant irony and outlandish yet human behaviors that have graced the body of Burton's work.<p>Augmented by deviously energetic performances from Glenn Shadix, Jeffrey Jones, Winona Ryder, Catherine O' Hara, Geena Davis, and Alec Baldwin, the film bustles with uninhibited brilliant hilarity that persistently pushes the film to a level of almost affable euphoria. Paired with its ingenious screenplay and a tour de force performance by Michael Keaton as Beetlejuice. This film transforms itself into an exuberant jovial exercise that is extremely satisfying and philosophically perceptive on many levels. Though comedy is usually considered to be unable or undeserving of deep critical analysis, Beetlejuice is an undeniably inspired concept that has been flawlessly transferred to film. One of the most outstanding comedies of the 1980's. <p>As for the film's DVD package, the DVD includes the theatrical trailer, an isolated Danny Elfman musical track, and the choice to watch the film in anamorphic widescreen or Pan and Scan. Hopefully Beetlejuice will eventually be graced with a special edition. We can always hope.

I'm sick of scads of nasty toothpicks all over my counter when I'm tinting icings; the squeeze-bottle concept is one that is long overdue and exceptionally welcome. The colors are vibrant, do not have an offensive taste, and blend, in my opinion, more smoothly than the Wilton pastes. You'll need to experiment a bit to get hues right at first, but it's well worth it.

I THOUGHT THIS MOVIE WAS SO FUNNY, MICHAEL KEATON AS BEETLEJUICE WAS RIGHT ON. AND ALEC BALDWIN PLAYING A GEEKY HUSBAND IS PRICELESS. I LOVE THE DA-OO SCENCE. THE KIDS AND I REALLY ENJOY THIS ONE.

I THOUGHT THIS MOVIE WAS SO FUNNY, MICHAEL KEATON AS BEETLEJUICE WAS RIGHT ON. AND ALEC BALDWIN PLAYING A GEEKY HUSBAND IS PRICELESS. I LOVE THE DA-OO SCENCE. THE KIDS AND I REALLY ENJOY THIS ONE.

I THOUGHT THIS MOVIE WAS SO FUNNY, MICHAEL KEATON AS BEETLEJUICE WAS RIGHT ON. AND ALEC BALDWIN PLAYING A GEEKY HUSBAND IS PRICELESS. I LOVE THE DA-OO SCENCE. THE KIDS AND I REALLY ENJOY THIS ONE.

Many movies, have dealt with the figure of death and the dead, from the Good angel clarence in "It's a Wonderful Life," to Brad Pitt's abysmal turn as Death himself in "Meet Joe Black." However, Tim Burton has presented the world with the most interesting look at the dead with Beetlejuice. In the first moments of the movie, Barbara and Adam Maitland go to town, And die in the oddest possible way. They return to thier house, and find them selves confined to it's premises. Thier leaving sends them to a strange desert world Populated by multi-layered sandworms. Eventually, they find that they are dead, and must remain in their home for 128 years, which wouldn't be so bad if it wasn't for the family of NYC artisans that moves in and remodels the place. After consulting with thier "caseworker" Juno,(Sylvia Sidney at her chain-smoking best) they hire a bio-exorcist,Betelguese (pronouced "BeetleJuice," hence the title) who agrees to help. Unfortunately for the Maitlands, Beetlejuice is an extremely PO'ed demon who refuses to return to the underworld. The visuals and special effects of this movie are completely stunning, and Micheal Keaton gives a performance that must have been a blast to play. Beetlejuice is a beer swilling,cussing,moral-majority's worst nightmare with superpowers, "Beetlejuice" is one of Tim Burton's Best works, and something he should be proud of.

Many movies, have dealt with the figure of death and the dead, from the Good angel clarence in "It's a Wonderful Life," to Brad Pitt's abysmal turn as Death himself in "Meet Joe Black." However, Tim Burton has presented the world with the most interestng look at the dead with Beetlejuice. In the first moments of the movie, Barbara and Adam Maitland go to town, And die in the oddest possible way. They return to thier house, and find them selves confined to it's premises. Thier leaving sends them to a strange desert world Populated by multi-layered sandworms. Eventually, they find that they are dead, and must remain in their home for 128 years, which wouldn't be so bad if it wasn't for the family of NYC artisans that moves in and remodels the place. After consulting with thier "caseworker" Juno,(Sylvia Sidney at her chain-smoking best) they hire a bio-exorcist,Betelguese (pronouced "BeetleJuice," hence the title) who agrees to help. Unfortunately for the Maitlands, Beetlejuice is an extremely PO'ed demon who refuses to return to the underworld. The visuals and special effects of this movie are completely stunning, and Micheal Keaton gives a performance that must have been a blast to play. Beetlejuice is a beer swilling,cussing,moral-majority's worst nightmare with superpowers, "Beetlejuice" is one of Tim Burton's Best works, and something he should be proud of.

Many movies, have dealt with the figure of death and the dead, from the Good angel clarence in "It's a Wonderful Life," to Brad Pitt's abysmal turn as Death himself in "Meet Joe Black." However, Tim Burton has presented the world with the most interestng look at the dead with Beetlejuice. In the first moments of the movie, Barbara and Adam Maitland go to town, And die in the oddest possible way. They return to thier house, and find them selves confined to it's premises. Thier leaving sends them to a strange desert world Populated by multi-layered sandworms. Eventually, they find that they are dead, and must remain in their home for 128 years, which wouldn't be so bad if it wasn't for the family of NYC artisans that moves in and remodels the place. After consulting with thier "caseworker" Juno,(Sylvia Sidney at her chain-smoking best) they hire a bio-exorcist,Betelguese (pronouced "BeetleJuice," hence the title) who agrees to help. Unfortunately for the Maitlands, Beetlejuice is an extremely PO'ed demon who refuses to return to the underworld. The visuals and special effects of this movie are completely stunning, and Micheal Keaton gives a performance that must have been a blast to play. Beetlejuice is a beer swilling,cussing,moral-majority's worst nightmare with superpowers, "Beetlejuice" is one of Tim Burton's Best works, and something he should be proud of.

I don't know why anyone would ever use those little liquid food colors again after trying these pastes. The colors are more intense, and you can easily blend lovely shades for your icings. <p>Martha Stewart shows wonderful seasonal cookies from time to time on her TV show. A simple eggwhite-sugar icing on cookies in exciting colors makes such a wonderful table decoration as well as great dessert for parties, teas and other events. These paste colors are a must if you like to make frosted cookies or party cakes. <p>And these are especially fun if you have kids. A cookie-icing contest is a fun event for a birthday party, Christmas party or just a rainy afternoon.

This food coloring kit sounds great except that they do not give the ounce sizes of the bottles. What's up with that? It would be nice to know how much product you are getting for your dollar.

Michael Keaton brings no distinguishing characteristics to the ghoul 'Beetlejuice', he merely acts bizarre, as does the script. It is often stunning cinematography but when the movie itself comes into focus, it's like finding one of Beetlejuice's snacks in your popcorn.

I am continually amazed at the shoddy treatment that some movies get in their DVD release. This DVD is simply a disgrace, especially considering what a great movie this is. I give the movie itself 5 stars; it's a wonderful example of Tim Burton's energy and style.<p>This DVD has no extras worth mentioning. No deleted scenes, no featurettes, not even a lousy commentary track! To make it even worse, the film has been CUT DOWN from the theatrical release! I have never seen a DVD release before where you get LESS than was originally presented in theaters.<p>My advice is to save your money until somebody figures out that when a movie is released on DVD, it needs to live up to the capabilities of the medium, and should always provide more material than was originally released, not less.

I am continually amazed at the shoddy treatment that some movies get in their DVD release. This DVD is simply a disgrace, especially considering what a great movie this is. I give the movie itself 5 stars; it's a wonderful example of Tim Burton's energy and style.<p>This DVD has no extras worth mentioning. No deleted scenes, no featurettes, not even a lousy commentary track! To make it even worse, the film has been CUT DOWN from the theatrical release! I have never seen a DVD release before where you get LESS than was originally presented in theaters.<p>My advice is to save your money until somebody figures out that when a movie is released on DVD, it needs to live up to the capabilities of the medium, and should always provide more material than was originally released, not less.

I am continually amazed at the shoddy treatment that some movies get in their DVD release. This DVD is simply a disgrace, especially considering what a great movie this is. I give the movie itself 5 stars; it's a wonderful example of Tim Burton's energy and style.<p>This DVD has no extras worth mentioning. No deleted scenes, no featurettes, not even a lousy commentary track! To make it even worse, the film has been CUT DOWN from the theatrical release! I have never seen a DVD release before where you get LESS than was originally presented in theaters.<p>My advice is to save your money until somebody figures out that when a movie is released on DVD, it needs to live up to the capabilities of the medium, and should always provide more material than was originally released, not less.

Just to warn you. when in trying to trick you in the widescreen format, the company made a mistake by also including the full screen version with it. i compared the widescreen and the full screen. and the full screen seems to contain more picture at the top and bottom. which must mean, that all they did was take the full screen version, cut off the top a bottom slightly and call it widescreen. which is strange because i would expect it to be easier just to make the theatrical version than to carefully edit the full screen and rip many people off.

Just to warn you. when in trying to trick you in the widescreen format, the company made a mistake by also including the full screen version with it. i compared the widescreen and the full screen. and the full screen seems to contain more picture at the top and bottom. which must mean, that all they did was take the full screen version, cut off the top a bottom slightly and call it widescreen. which is strange because i would expect it to be easier just to make the theatrical version than to carefully edit the full screen and rip many people off.

Just to warn you. when in trying to trick you in the widescreen format, the company made a mistake by also including the full screen version with it. i compared the widescreen and the full screen. and the full screen seems to contain more picture at the top and bottom. which must mean, that all they did was take the full screen version, cut off the top a bottom slightly and call it widescreen. which is strange because i would expect it to be easier just to make the theatrical version than to carefully edit the full screen and rip many people off.

I bought these to decorate some dia de los muertos skulls with icing. The Ateco food coloring gave nice bright, vibrant colors and mixed with the icing very well. I was totally satisfied with the purchase.

Winona Ryder is HOT! HOT! HOT! as the gothic death-rock princess of Doom (see "Lucas" for more W.R. action) The best performance from that eldest Baldwin brother that I've ever seen from any of them.His zenith I'm sure. BEST PERFORMANCE - goes to - Miss Ryder!,of course, I'm a BIG fan, you see? FAV PART - when Beetlejuice was chumming up to Baldwin's character while wearing the same type clothes. No big deal, just witty and clever charm with spit. and when Beetlejuice hits up on Gina Davis and/or Winona Ryder

Winona Ryder is HOT! HOT! HOT! as the gothic death-rock princess of Doom (see "Lucas" for more W.R. action) The best performance from that eldest Baldwin brother that I've ever seen from any of them.His zenith I'm sure. BEST PERFORMANCE - goes to - Miss Ryder!,of course, I'm a BIG fan, you see? FAV PART - when Beetlejuice was chumming up to Baldwin's character while wearing the same type clothes. No big deal, just witty and clever charm with spit. and when Beetlejuice hits up on Gina Davis and/or Winona Ryder

```
In [27]: %%sql
--Check for Missing Reviews
SELECT COUNT(*) AS missing_reviews
FROM Reviews
WHERE Text IS NULL OR TRIM(Text) = '';
```

```
* sqlite:///database.sqlite
Done.
```

```
Out[27]: missing_reviews
         0
```

```
In [28]: %%sql

SELECT
    CASE
        WHEN LENGTH(Text) < 50 THEN 'Very Short'
        WHEN LENGTH(Text) BETWEEN 50 AND 150 THEN 'Short'
        WHEN LENGTH(Text) BETWEEN 151 AND 300 THEN 'Medium'
        ELSE 'Long'
    END AS review_length_category,
    COUNT(*) AS count
FROM Reviews
GROUP BY review_length_category
ORDER BY count DESC
LIMIT 100;
```

```
* sqlite:///database.sqlite
Done.
```

```
Out[28]: review_length_category  count
         Long      285185
         Medium    186686
         Short     96512
         Very Short      71
```

The analysis of review length distribution provides valuable insights into customer engagement and communication patterns. By categorizing reviews into length groups (Very Short, Short, Medium, and Long), it's evident that the majority of reviews fall into the "Short" and "Medium" categories, suggesting that customers generally prefer to provide concise feedback. This information can be useful for optimizing platforms and interfaces to encourage efficient review submissions. Additionally, identifying a subset of "Long" reviews can highlight customers who are highly engaged and provide detailed feedback, offering opportunities for in-depth analysis to extract specific product strengths and weaknesses. Understanding these patterns allows businesses to tailor their strategies for collecting and utilizing customer feedback more effectively.

```
In [29]: %%sql
--Find Peak Review Activity
SELECT Time, COUNT(*) AS count
FROM Reviews
GROUP BY Time
ORDER BY count DESC
LIMIT 20;
```

```
* sqlite:///database.sqlite
Done.
```

Out[29]:

Time	count
1350345600	1143
1322179200	1088
1322438400	1070
1346889600	1018
1344211200	989
1323216000	978
1350518400	915
1334793600	906
1346976000	901
1347062400	898
1332374400	898
1349740800	895
1332892800	890
1349654400	878
1350432000	877
1325203200	872
1349827200	870
1325808000	867
1347235200	866
1350864000	863

In [30]:

```
%%sql
SELECT
    Time,
    DATETIME(Time, 'unixepoch') AS readable_time
FROM Reviews
LIMIT 10;
```

* sqlite:///database.sqlite
Done.

Out[30]:

Time	readable_time
1303862400	2011-04-27 00:00:00
1346976000	2012-09-07 00:00:00
1219017600	2008-08-18 00:00:00
1307923200	2011-06-13 00:00:00
1350777600	2012-10-21 00:00:00
1342051200	2012-07-12 00:00:00
1340150400	2012-06-20 00:00:00
1336003200	2012-05-03 00:00:00
1322006400	2011-11-23 00:00:00
1351209600	2012-10-26 00:00:00

In [31]:

```
%%sql
-- Find Peak Review Activity with Readable Time
SELECT
    Time,
    DATETIME(Time, 'unixepoch') AS readable_time,
    COUNT(*) AS count
FROM Reviews
GROUP BY Time
ORDER BY count DESC
LIMIT 20;
```

* sqlite:///database.sqlite
Done.

Out[31]:	Time	readable_time	count
	1350345600	2012-10-16 00:00:00	1143
	1322179200	2011-11-25 00:00:00	1088
	1322438400	2011-11-28 00:00:00	1070
	1346889600	2012-09-06 00:00:00	1018
	1344211200	2012-08-06 00:00:00	989
	1323216000	2011-12-07 00:00:00	978
	1350518400	2012-10-18 00:00:00	915
	1334793600	2012-04-19 00:00:00	906
	1346976000	2012-09-07 00:00:00	901
	1347062400	2012-09-08 00:00:00	898
	1332374400	2012-03-22 00:00:00	898
	1349740800	2012-10-09 00:00:00	895
	1332892800	2012-03-28 00:00:00	890
	1349654400	2012-10-08 00:00:00	878
	1350432000	2012-10-17 00:00:00	877
	1325203200	2011-12-30 00:00:00	872
	1349827200	2012-10-10 00:00:00	870
	1325808000	2012-01-06 00:00:00	867
	1347235200	2012-09-10 00:00:00	866
	1350864000	2012-10-22 00:00:00	863

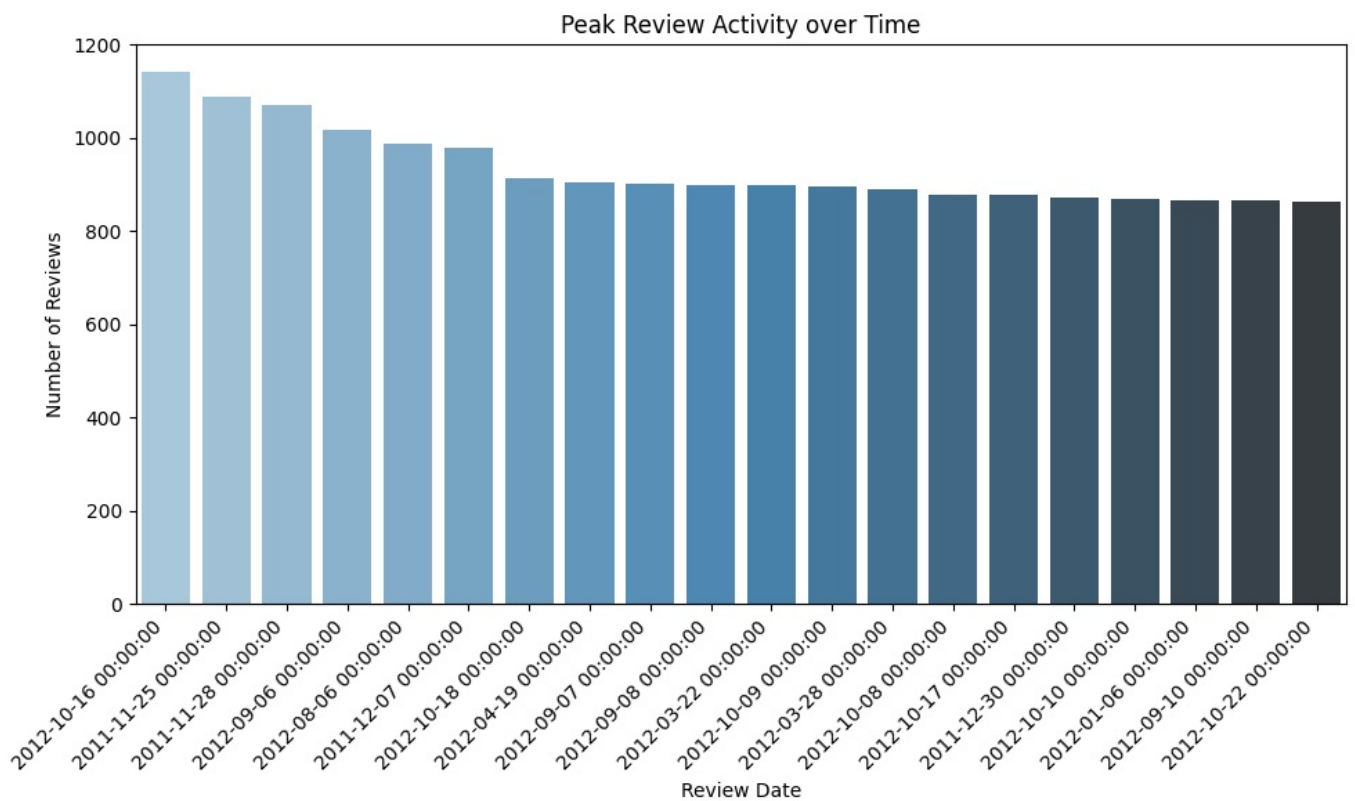
```
In [32]: import warnings
warnings.filterwarnings('ignore')
```

```
In [37]: import sqlite3
import importlib
import algorithms
from algorithms import plot_peak_review_activity #Algorithm to plot peak review activity
# using the above Peak review Activity SQL Query

# Connect to the database
conn = sqlite3.connect("database.sqlite")

# Plot the review activity
plot_peak_review_activity()

# Close the connection
conn.close()
```



```
In [1]: import sqlite3
import matplotlib.pyplot as plt
import nltk
from algorithms import analyze_reviews, generate_wordcloud # Import functions

# Ensure that necessary NLTK resources are downloaded (do this once)
nltk.download('punkt')
nltk.download('stopwords')

# Connect to the SQLite database
conn = sqlite3.connect('database.sqlite')

# Analyze reviews using the function from algorithms.py
word_frequency_dict = analyze_reviews(conn)

# Generate the word cloud
wordcloud = generate_wordcloud(word_frequency_dict)

# Plot the WordCloud
plt.figure(figsize=(10, 5))
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis('off')
plt.title('Top 200 Words Word Cloud', fontsize=20)
plt.show()

# Close the database connection
conn.close()
```

```
[nltk_data] Downloading package punkt to C:\Users\Collins
[nltk_data]   PC\AppData\Roaming\nltk_data...
[nltk_data]   Package punkt is already up-to-date!
[nltk_data] Downloading package stopwords to C:\Users\Collins
[nltk_data]   PC\AppData\Roaming\nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
```


[illegible]

```
* sqlite:///database.sqlite
Done.
```

A3OXHLG6DIBRW8	448
A1YUL9PCJR3JTY	421
AY12DBB0U420B	389
A281NPSIMI1C2R	365
A1Z54EM24Y40LL	256
A1TMAVN4CEM8U8	204
A2MUGFV2TDQ47K	201
A3TVZM3ZIXG8YW	199
A3PJZ8T8U8FDQ1K	178
AQQLWCMRNDFGI	176

```
* sqlite:///database.sqlite
Done.
```


Out[39]:

Year	ReviewCount
------	-------------

1001	3
1003	1
1004	2
1009	1
1010	3
1012	3
1013	6
1014	6
1015	6
1018	1
1020	1
1022	1
1023	6
1024	3
1025	4
1029	1
1032	1
1035	1
1036	17
1037	6

```
In [40]: %%sql
--Most Reviewed Products
SELECT ProductId, COUNT(*) AS ReviewCount
FROM Reviews
GROUP BY ProductId
ORDER BY ReviewCount DESC
LIMIT 10;
```

* sqlite:///database.sqlite
Done.

Out[40]:

ProductId	ReviewCount
-----------	-------------

B007JFMH8M	913
B002QWP8H0	632
B002QWP89S	632
B002QWHJOU	632
B0026RQTGE	632
B003B3OOPA	623
B001EO5Q64	567
B007M83302	564
B007M832YY	564
B006HYLW32	564

```
In [69]: import sqlite3
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Connect to SQLite database
conn = sqlite3.connect('database.sqlite')

# SQL query to get the top 10 most reviewed products
query = """
SELECT ProductId, COUNT(*) AS ReviewCount
```

```

FROM Reviews
GROUP BY ProductId
ORDER BY ReviewCount DESC
LIMIT 10;
"""

# Load data into a Pandas DataFrame
df = pd.read_sql(query, conn)

# Close the database connection
conn.close()

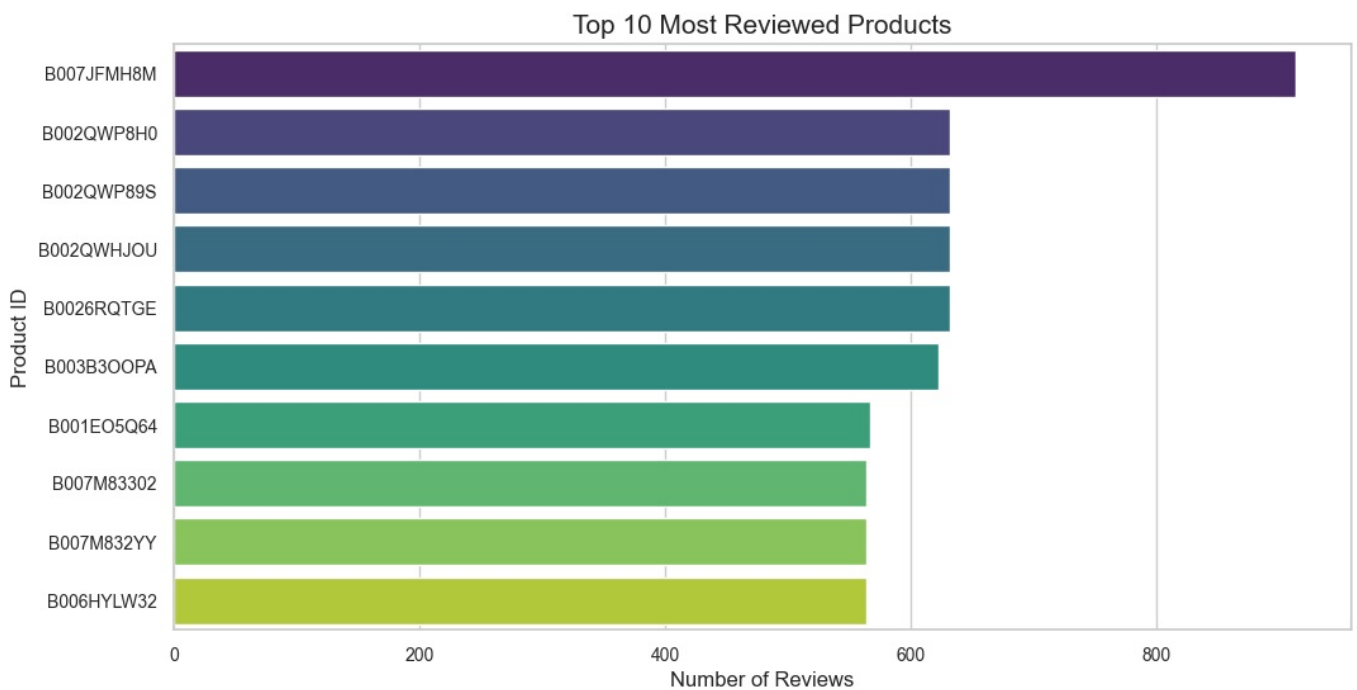
# Set Seaborn style
sns.set_theme(style="whitegrid")

# Create a bar plot
plt.figure(figsize=(12, 6))
sns.barplot(x=df["ReviewCount"], y=df["ProductId"], palette="viridis")

# Add labels and title
plt.xlabel("Number of Reviews", fontsize=12)
plt.ylabel("Product ID", fontsize=12)
plt.title("Top 10 Most Reviewed Products", fontsize=15)
plt.xticks(fontsize=10)
plt.yticks(fontsize=10)

# Show the plot
plt.show()

```



```

In [22]: import warnings
warnings.filterwarnings('ignore')

```

```

In [23]: %%sql
--Average Score by product
SELECT ProductId, AVG(Score) AS AvgScore
FROM Reviews
GROUP BY ProductId
ORDER BY AvgScore DESC
LIMIT 30;

* sqlite:///database.sqlite
Done.

```

Out[23]:

ProductId	AvgScore
B009WVB40S	5.0
B009WSNWC4	5.0
B009UUS05I	5.0
B009SR4OQ2	5.0
B009SF0TN6	5.0
B009RSR8HO	5.0
B009RE0Y5G	5.0
B009PIEW3O	5.0
B009PICJTS	5.0
B009PIAFTE	5.0
B009PG8MVO	5.0
B009PFJUF2	5.0
B009PCDDO4	5.0
B009OY38SY	5.0
B009OM66IU	5.0
B009OM65H2	5.0
B009OM65GI	5.0
B009O7DGEW	5.0
B009O753WA	5.0
B009O0U2SW	5.0
B009NWIF3A	5.0
B009NIF7BM	5.0
B009MQVNW2	5.0
B009M6W5UQ	5.0
B009LT26BC	5.0
B009LSMTWE	5.0
B009LIK0FW	5.0
B009L8MUD2	5.0
B009L5LIVA	5.0
B009L47BR6	5.0

In [24]:

```
import sqlite3
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Connect to SQLite database
conn = sqlite3.connect('database.sqlite')

# SQL query to get the top 30 products with the highest average score
query = """
SELECT ProductId, AVG(Score) AS AvgScore
FROM Reviews
GROUP BY ProductId
ORDER BY AvgScore DESC
LIMIT 30;
"""

# Load data into a Pandas DataFrame
df = pd.read_sql(query, conn)

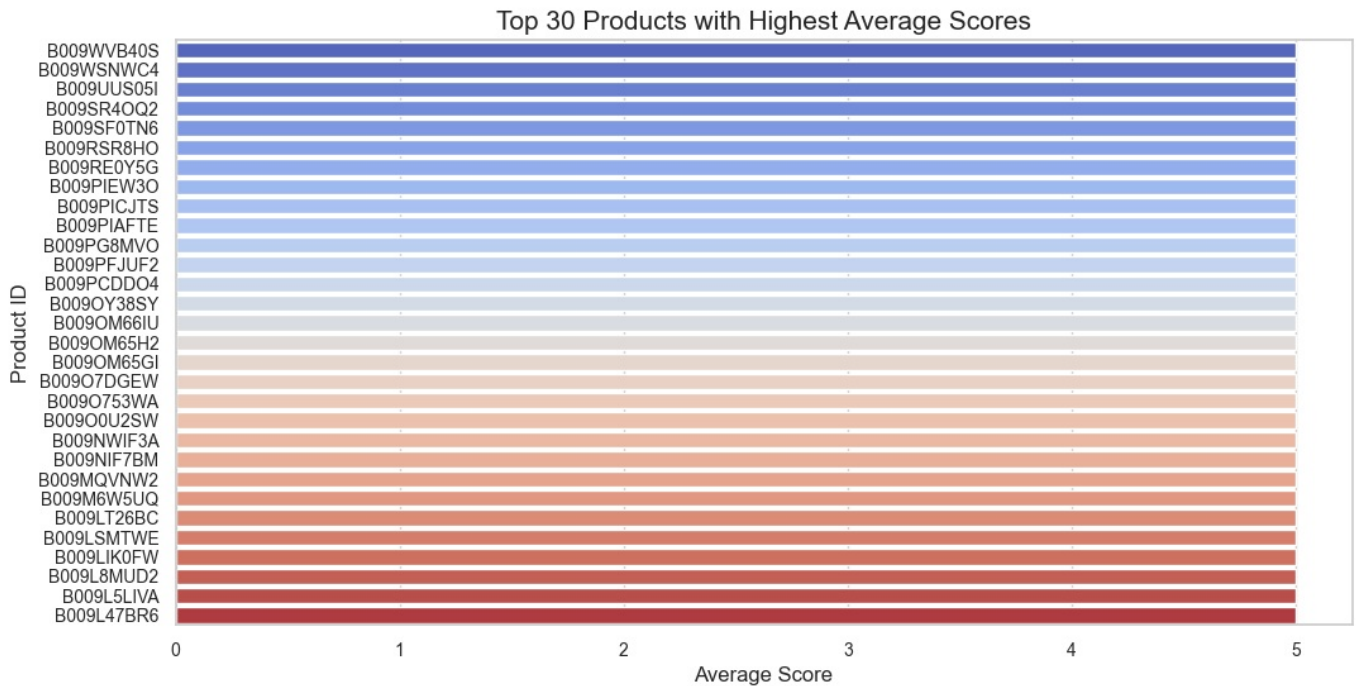
# Close the database connection
conn.close()

# Set Seaborn style
sns.set_theme(style="whitegrid")

# Create a bar plot
plt.figure(figsize=(12, 6))
sns.barplot(x=df["AvgScore"], y=df["ProductId"], palette="coolwarm")
```

```
# Add labels and title
plt.xlabel("Average Score", fontsize=12)
plt.ylabel("Product ID", fontsize=12)
plt.title("Top 30 Products with Highest Average Scores", fontsize=15)
plt.xticks(fontsize=10)
plt.yticks(fontsize=10)

# Show the plot
plt.show()
```



```
In [25]: %%sql
--Most Frequent Word Summaries for Different Products
SELECT ProductId, Summary, COUNT(*) AS Frequency
FROM Reviews
WHERE Summary IS NOT NULL AND Summary <> ''
GROUP BY ProductId, Summary
ORDER BY Frequency DESC
LIMIT 10;
```

```
* sqlite:///database.sqlite
Done.
```

```
Out[25]:
```

ProductId	Summary	Frequency
B0026RQTGE	Greenies	30
B002QWHJOU	Greenies	30
B002QWP89S	Greenies	30
B002QWP8H0	Greenies	30
B007JFMH8M	Yummy	22
B007JFMH8M	Yummy!	21
B003B3OOPA	Coconut Oil	18
B007JFMH8M	Quaker Soft Baked Oatmeal Cookies	18
B007JFMH8M	Delicious	16
B007JFMH8M	Delicious!	16

```
In [26]: import sqlite3
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Connect to SQLite database
conn = sqlite3.connect('database.sqlite')

# SQL query to get the top 10 most frequent review summaries
query = """
SELECT Summary, COUNT(*) AS Frequency
FROM Reviews
GROUP BY Summary
ORDER BY Frequency DESC
LIMIT 10;
"""
```

```

# Load data into a Pandas DataFrame
df = pd.read_sql(query, conn)

# Close the database connection
conn.close()

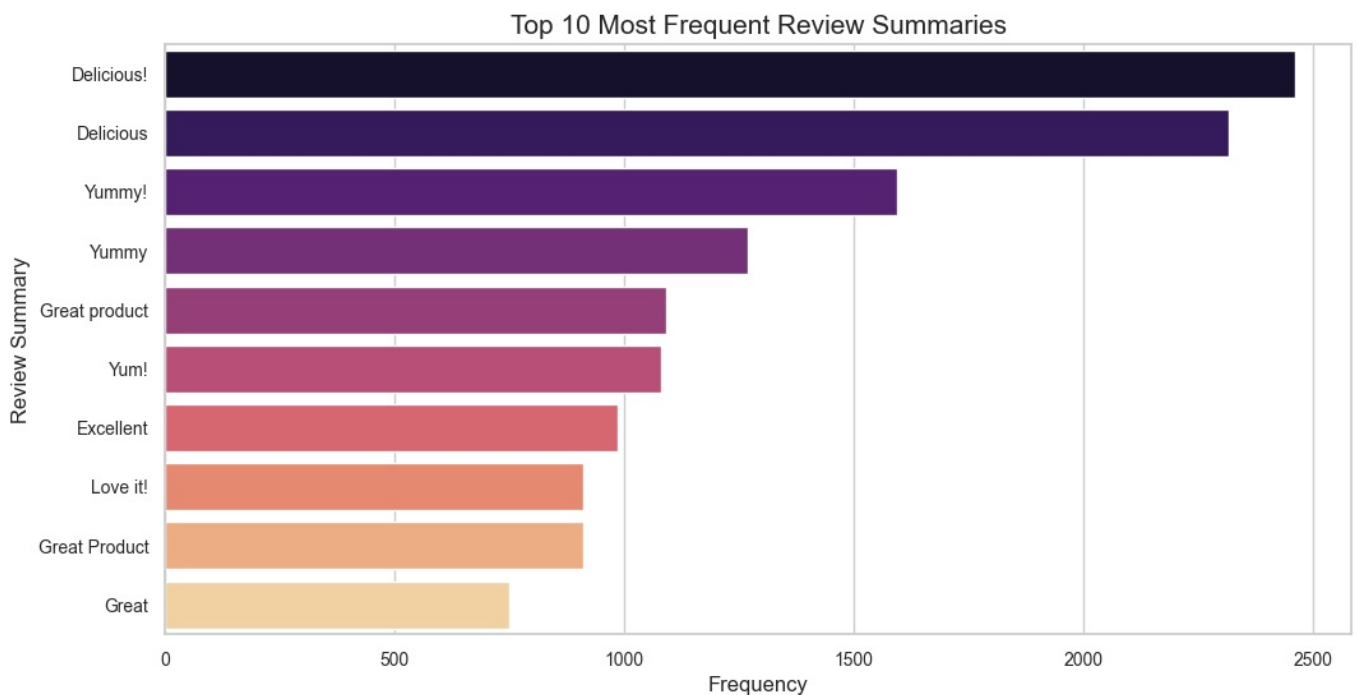
# Set Seaborn style
sns.set_theme(style="whitegrid")

# Create a bar plot
plt.figure(figsize=(12, 6))
sns.barplot(x=df["Frequency"], y=df["Summary"], palette="magma")

# Add labels and title
plt.xlabel("Frequency", fontsize=12)
plt.ylabel("Review Summary", fontsize=12)
plt.title("Top 10 Most Frequent Review Summaries", fontsize=15)
plt.xticks(fontsize=10)
plt.yticks(fontsize=10)

# Show the plot
plt.show()

```



NLP Analysis

```

In [41]: import sqlite3
import pandas as pd

# Connect to SQLite database
conn = sqlite3.connect("database.sqlite")

# Read entire Reviews table into a Pandas DataFrame
reviews = pd.read_sql("SELECT * FROM Reviews", conn)

# Close the connection
conn.close()

# Display the first few rows
print(reviews.head())

```

	Id	ProductId	UserId	ProfileName	\
0	1	B001E4KFG0	A3SGXH7AUHU8GW	delmartian	
1	2	B00813GRG4	A1D87F6ZCVE5NK	dll pa	
2	3	B000LQOCH0	ABXLMWJIXXAIN	Natalia Corres	"Natalia Corres"
3	4	B000UA0QIQ	A395BORC6FGVXV	Karl	
4	5	B006K2ZZ7K	A1UQRSCLF8GW1T	Michael D. Bigham	"M. Wassir"

	HelpfulnessNumerator	HelpfulnessDenominator	Score	Time	\
0	1	1	5	1303862400	
1	0	0	1	1346976000	
2	1	1	4	1219017600	
3	3	3	2	1307923200	
4	0	0	5	1350777600	

	Summary	Text
0	Good Quality Dog Food	I have bought several of the Vitality canned d...
1	Not as Advertised	Product arrived labeled as Jumbo Salted Peanut...
2	"Delight" says it all	This is a confection that has been around a fe...
3	Cough Medicine	If you are looking for the secret ingredient i...
4	Great taffy	Great taffy at a great price. There was a wid...

In [42]: reviews

Out[42]:

	Id	ProductId	UserId	ProfileName	HelpfulnessNumerator	HelpfulnessDenominator	Score	Time
--	----	-----------	--------	-------------	----------------------	------------------------	-------	------

0	1	B001E4KFG0	A3SGXH7AUHU8GW	delmartian	1	1	5	1303862400
1	2	B00813GRG4	A1D87F6ZCVE5NK	dll pa	0	0	1	1346976000
2	3	B000LQOCH0	ABXLMWJIXXAIN	Natalia Corres "Natalia Corres"	1	1	4	1219017600
3	4	B000UA0QIQ	A395BORC6FGVXV	Karl	3	3	2	1307923200
4	5	B006K2ZZ7K	A1UQRSCLF8GW1T	Michael D. Bigham "M. Wassir"	0	0	5	1350777600
...
568449	568450	B001EO7N10	A28KG5XORO54AY	Lettie D. Carter	0	0	5	1299628800
568450	568451	B003S1WTCU	A3I8AFVPEE8KI5	R. Sawyer	0	0	2	1331251200
568451	568452	B004I613EE	A121AA1GQV751Z	pkasd "pk_007"	2	2	5	1329782400
568452	568453	B004I613EE	A3IBEVCTXKNOH	Kathy A. Welch "katwel"	1	1	5	1331596800
568453	568454	B001LR2CU2	A3LGQPJCZVL9UC	srfell17	0	0	5	1338422400

568454 rows × 10 columns

◀ ▶

In [43]: reviews.describe()

Out[43]:

	Id	HelpfulnessNumerator	HelpfulnessDenominator	Score	Time
count	568454.000000	568454.000000	568454.00000	568454.000000	5.684540e+05
mean	284227.500000	1.743817	2.22881	4.183199	1.296257e+09
std	164098.679298	7.636513	8.28974	1.310436	4.804331e+07
min	1.000000	0.000000	0.00000	1.000000	9.393408e+08
25%	142114.250000	0.000000	0.00000	4.000000	1.271290e+09
50%	284227.500000	0.000000	1.00000	5.000000	1.311120e+09
75%	426340.750000	2.000000	2.00000	5.000000	1.332720e+09
max	568454.000000	866.000000	923.00000	5.000000	1.351210e+09

In [44]: reviews['Summary'].value_counts()

Out[44]: Summary
Delicious! 2462
Delicious 2316
Yummy! 1594
Yummy 1268
Great product 1091
...
Makes a super easy jerk marinade 1
Best for Diabetics 1
A great change for those who love rice. 1
white rice 1
Favorite Training and reward treat 1
Name: count, Length: 295744, dtype: int64

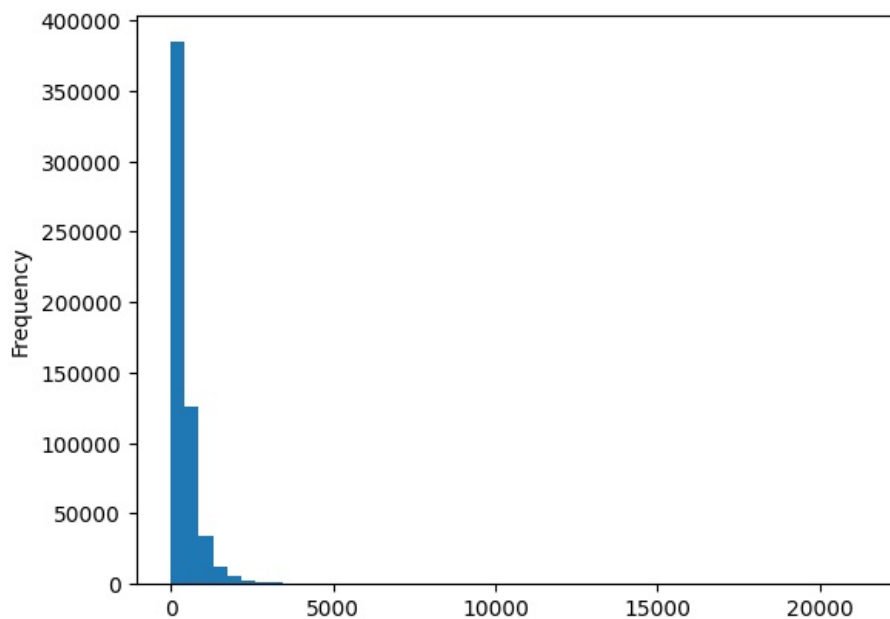
In [45]: reviews['length'] = reviews['Text'].apply(len)

In [46]: reviews

	Id	ProductId	UserId	ProfileName	HelpfulnessNumerator	HelpfulnessDenominator	Score	Time
0	1	B001E4KFG0	A3SGXH7AUHU8GW	delmartian	1	1	5	130386240
1	2	B00813GRG4	A1D87F6ZCVE5NK	dll pa	0	0	1	134697600
2	3	B000LQOCH0	ABXLMWJIXXAIN	Natalia Corres "Natalia Corres"	1	1	4	121901760
3	4	B000UA0QIQ	A395BORC6FGVXV	Karl	3	3	2	130792320
4	5	B006K2ZZ7K	A1UQRSCLF8GW1T	Michael D. Bigham "M. Wassir"	0	0	5	135077760
...
568449	568450	B001EO7N10	A28KG5XORO54AY	Lettie D. Carter	0	0	5	129962880
568450	568451	B003S1WTCU	A3I8AFVP EE8KI5	R. Sawyer	0	0	2	133125120
568451	568452	B004I613EE	A121AA1GQV751Z	pk sd "pk_007"	2	2	5	132978240
568452	568453	B004I613EE	A3IBEVCTXKNOH	Kathy A. Welch "katwel"	1	1	5	133159680
568453	568454	B001LR2CU2	A3LGQPJCZVL9UC	srfell17	0	0	5	133842240

```
reviews['length'].plot.hist(bins=50)
```

```
<Axes: ylabel='Frequency'>
```

```
In [48]: reviews['length'].describe()
```

```
Out[48]: count    568454.000000
         mean      436.222083
         std       445.339741
         min       12.000000
         25%      179.000000
         50%      302.000000
         75%      527.000000
         max      21409.000000
         Name: length, dtype: float64
```

```
In [49]: reviews[reviews['length']==21409]['Text'].iloc[0]
```

```
Out[49]: 'Fuzzy Wuzzy\'s Summary:<br />*** Somewhat recommended, with reservations and only lukewarm fuzzies.<br /><br />
Positives:<br /><br />+ Not too expensive compared to other bottled water, although it is not cheap either.<br />
Negatives:<br /><br />- This bottled water product\'s labeling, product descriptions, and health claims
are misleading.<br /><br />At first glance, this 1.5-liter bottle of water that I am reviewing looks like another
one of the many varieties of bottled water on the market. There are many kinds of bottled water: "spring water"
produced from an underground formation where the groundwater naturally flows up to the Earth\'s surface, "
artesian water" obtained from a well that taps into an underground aquifer, "mineral water" produced from a well
or spring that naturally contains trace amounts of minerals that some people consider to provide health benefits,
"ground/well water" produced from a well that taps into an aquifer, "sparkling water" from a spring or well that
is naturally carbonated, and "purified water" produced by distillation, deionization, reverse osmosis, or other
purification processes. This Essentia Drinking Water is "purified water"; this bottle contains municipal water
(i.e. tap water that comes from your faucet) that is purified using reverse osmosis and ultraviolet light for
disinfection.<br /><br />The label on this Essentia water bottle is filled with marketing hype and technical terms
designed to make it sound exotic:<br />* "Infused with Electrolytes For A Difference You Can Taste"
<br />* "Silky Smooth Taste"
<br />* "Hydrates Better"
<br />* "Electrolytes Infused"
<br />* "Best Available pH at 9.5"
<br />* "Doctor recommended to protect, improve, and enhance the quality of your health."
<br />* "Our Proprietary Technology: Water purified by reverse osmosis; infused with electrolytes for taste; alkalized
and restructured using Ionic Separation Technology."
<br />* "Ingredients: purified water, sodium bicarbonate, dipotassium phosphate, magnesium sulfate, calcium chloride."
<br /><br />There are two ways to evaluate this water product. One way to review this water is based on how it tastes,
how safe are its added ingredients, and its cost. Since Essentia includes a variety of health claims and marketing
statements on the water bottle\'s label, on
```

the Amazon product listing, and on the company's Web site, the veracity of this statements must also be scrutinized since the marketing hype of most other brands of bottled water is just limited to the "freshness" or "purity" of their taste without making other health-related claims.

First, on the basis on its cost, with all of its marketing hype and hoopla, at least this water is not priced far more expensive than most other brands of bottled water. But there are various other bottled water products on Amazon that cost less on a per-liter basis. And if you think that this slightly pricey water can be justified based on its health "merits", keep reading...

As for taste, that can be a highly personal preference, but I like my water to simply taste "wet". :-) If I want some kind of flavored drink - tea, wine, beer, fruit juice, etc. - I will drink that instead. But nothing hydrates and quenches as easily as pure and simple water, with no additives or flavorings added to it. This particular Essentia water does have a smooth texture to it. Most other bottled waters taste more crisp than this Essentia water. Again, you may either like the smooth texture of this water or you may think that its texture is rather flat and bland. Unlike drinking "distilled water", or bottled water that was purified using distillation with no additives, this water has a very slight subtle aftertaste that might be considered as either "chalky" or "metallic". Those are my highly subjective adjectives that I associate with this water. I can only surmise that this is due to the small amounts of additives that were used in this water. This aftertaste is not too objectionable, but if I just want to hydrate myself drinking "water", I really prefer my water to be without any identifiable taste, texture, or odor.

Next, since this is not just a bottle of pure H₂O water, let's consider what was added to this water: sodium bicarbonate, dipotassium phosphate, magnesium sulfate, and calcium chloride. Sodium bicarbonate is the same thing as baking soda. This water was made to have a slightly alkaline pH of 9.5 by adding sodium bicarbonate. People sometimes also adjust and raise the pH of their swimming pool water by adding sodium bicarbonate. Dipotassium phosphate is sometimes added to various food products as a buffering agent or emulsifier. In this water, it serves as one of the "electrolytes" that is added. The term "electrolytes" is just a subset of minerals that are found in both the Earth and the food that we eat. Also serving as an electrolyte is the addition of magnesium sulfate. Magnesium sulfate is sold in drug stores as "Epsom Salt", and you can either soak your feet or body in an Epsom Salt bath to soothe sore muscles, or when ingested in larger quantities, magnesium sulfate also can serve as a laxative. Calcium chloride serves as another electrolyte in this water. Calcium chloride is sometimes used as a food additive. The trace amounts of calcium, magnesium, and phosphorus in this water do not qualify as a measurable part of your daily nutritional intake of these minerals, regardless of how much of this water you drink each day.

Essentia is definitely not the first bottled water manufacturer to add electrolytes to either bottled water or drinks. During the 1960s, Gatorade was possibly the first large-scale commercial drink that was produced with the goal of adding electrolytes of sodium, sugar, potassium, and phosphate to hydrate and replenish athletes who lost fluids during physical exertion in hot weather. On Essentia's Web site, they have an FAQ question/answer of: "Why should I drink Essentia if I get my electrolytes through sports drinks? Essentia provides a blend of dissolved minerals that emulate the ratio of the body's own mixture and concentration - providing the correct proportion of electrolytes without all the sugars and other additives you typically find in sports and energy drinks." There are several problems with that statement. First, if you are actively exercising and sweating, your body can actually use those extra calories, and if you are just sitting in front of the computer or television, guzzling Gatorade just adds useless calories and electrolytes. Second, if you are drinking Essentia at home, this product's claim that it provides "the correct proportion of electrolytes" implies a one-size-fits-all balance of electrolytes, which is simply not true as we all have different bodies, diets, and lifestyles.

None of what Essentia refers to as "Our Proprietary Technology" is really that "proprietary" or exotic. The reverse osmosis purification of municipal tap water is commonly used to produce bottled water, and the addition of various electrolytes or minerals to water has been practiced by other bottled water manufacturers for many decades now. Some people think that drinking plain distilled water tastes boring and, mentally at least, they like the idea that drinking water that is naturally infused with minerals from a spring or well/aquifer adds an extra measure of health benefits. The addition of trace minerals do add noticeable flavor that mimics natural spring water. Also, slightly flavoring the water with minerals masks the residual odors that can be left by the plastic water bottle; the plastic odor is still there, but you notice it less. The same principle applies to using a room deodorizer to mask the existing odors in a house; that old cat urine smell is still present inside the house, but you notice it less and therefore the house smells fresher. The most common electrolytes/minerals in spring water are calcium, magnesium, potassium, sodium, sulfate, and chloride. The Coca-Cola company markets a "Dasani" bottled water that, like this Essentia, is municipal tap water purified via reverse osmosis. Dasani has magnesium sulfate and potassium chloride added to it. Perhaps the Dasani taste testing panels decided that their water tastes better without the addition of calcium? But for whatever reasons, this Essentia water has a slight residual "chalky" or "metallic" aftertaste, as if it had a tannic undertone like the aftertaste after drinking a dry red wine, that I do not really care for. I do not want my drinking water to leave a subtle aftertaste in my mouth.

Essentia also competes directly with "Glaceau Smart Water", which takes an approach similar to Essentia's water. Glaceau Smart Water is municipal water purified via distillation, and then infused with the similar ingredients of calcium chloride, magnesium chloride, and potassium bicarbonate.

Not to be outdone, another of Essentia's alkaline/electrolyte water competitors, named "NEO Water" (to ride on the futuristic coattails of "The Matrix" fame??), pumps out even more junk science driven filled with fear-mongering misleading statements, calling itself "Alkaline, Electrolyte, Antioxidant Super Hydrator ... ultra purified healthy and refreshing water formulated through proprietary technology to balance, revitalize and propel you unlike any traditional water", and "NEO's unique antioxidants neutralize free radicals that the body releases when under exertion". But when you really dig further into NEO Water's claims, there are ***NO*** antioxidants in NEO Water... just potassium bicarbonate, calcium chloride, magnesium chloride, which is similar to the mundane chemicals added to Essentia and Dasani water. NEO Water's marketing nonsense says that its "alkalinity allows its molecules to carry more oxygen than normal with an extra attached electron (hydroxyl ions). As a result, these molecules are able to seek and neutralize free radicals in your body and prevent free radical damage to your cells. These natural antioxidant benefits leave your body rich in oxygen and energy." Again, like Essentia, NEO Water hopes to blind and persuade the consumer by overwhelming you with exotic totally false techno-babble; it ascribes its antioxidant benefits entirely to its 9.5 pH alkalinity. So if you really believe in this nonsense, you could, in theory, obtain far more amazing antioxidant protection simply by dissolving an antacid tablet into your tap water once a day. As with Essentia's water, NEO Water's "ultra purified" water is simply municipal tap water purified using the common method of reverse osmosis.

This Essentia Drinking Water is not the first product to tout the benefits of "alkaline water". There have been previous highly-hyped products based on ionized water, alkaline water, and combination alkaline-ionized water. There have even been so-called "liquid oxygen" drops that supposedly contained "O₄" molecules (instead of the "O₂" oxygen that is in the air) dissolved in a liquid, and by ingesting 10 drops each day, you feel more energized, but this quackery completely ignores the fact that oxygen is best utilized in the lungs, and not in the stomach. Just as the "liquid oxygen" drops are a distorted version of twisted junk science based on the concept of oxygen bars (at least in oxygen bars, you do breathe in the oxygen instead of putting drops of useless oxygen in your

stomach), this "alkaline water" concept is a distortion of the "alkaline diet" alternative medicine belief that the foods we eat can affect the acidity of urine and blood, which can then affect our well-being, susceptibility to disease, and energy level. If you are an avid vegan/vegetarian practitioner of an alkaline diet, then this 9.5-pH alkaline water can keep your diet on its alkaline-based theme. But if you regularly eat meats, poultry, seafood, dairy products, eggs, grains, processed sugar foods, and colas/sodas, all of which significantly boost the acidic levels of your food consumption, then drinking a slightly alkaline water makes absolutely no difference to your health.

Your stomach has a healthily acidic pH that ranges between 1.45 to 3.5; your stomach must be very acidic to aid in digestion. However, blood must always be slightly alkaline, with a pH of 7.35 to 7.45. If the blood's pH drops below 6.8 (more acidic) or rises above 7.8 (more alkaline), death may occur. But your blood automatically stays within its healthy pH range regardless of what you eat or drink because your blood utilizes several buffering systems. The most important buffer for maintaining acidity/alkalinity balance in the blood is called the carbonic-acid-bicarbonate buffer. To form the bicarbonate in this buffering system, your body uses carbon dioxide... guess what, carbon dioxide is in (over)abundance in the air that you breathe and you do not need sodium bicarbonate baking soda added to your drinking water to maintain a healthy pH in your body. In fact, consuming too much supplemental sodium bicarbonate can actually lead to constipation, convulsions, diarrhea, frequent urination, and vomiting. There is only a tiny amount of sodium bicarbonate in each bottle of Essentia water, so you do not have to worry about overdose. But my point is that all this hoopla and hype about this water being of the "Best Available pH at 9.5" is purified nonsense. Some other fear-mongering alkaline water sales gimmicks claim that, to quote another Web site, "Excess acidity weakens all our systems and forces the body to take minerals from vital organs and bones to buffer/neutralize the acid and safely remove it from the body." But this is totally false fear-based junk science contrived to mislead you into thinking that your body will rob its vital organs and bones of minerals to neutralize excess acidity. But our body's carbonic-acid-bicarbonate buffer simply combines carbon dioxide (CO₂) in the air with water to make the blood plasma more acidic if it needs to, and it can prevent excess plasma acidity and raise its pH to be more alkaline by having the lungs convert the bicarbonate back into CO₂ which you then breathe out. A similar phosphate buffering system also helps to maintain the pH of cellular fluids, using phosphate instead of bicarbonate. Phosphorus is present in nearly all foods, and regardless of whether you like to eat lots of meats or lots of vegetables, you can get plenty of sufficient phosphorus in your diet without needing to add it to your drinking water.

Drinking too many acidic drinks such as sodas and colas can have a detrimental effect of eroding your tooth enamel, but guzzling all those colas have absolutely no effect on the pH of your cells and bodily fluids. And regardless of how much alkalinity you consume and put in your stomach, your body is going to make sure that your stomach stays within a healthy acidic pH range and your body will use its pH buffering systems to keep your blood pH between 7.35 and 7.45 regardless of how much acidic or alkaline food and drink you consume. If you are familiar with logarithms, also realize that the pH scale is logarithmic (pH is not a linear scale) and drinking 9.5 pH water really has no effect on your various bodily fluids compared with drinking tap water or any other bottled water. And if you really could somehow artificially adjust your stomach, blood, and bodily fluids entirely to a 9.5 pH, you would be immediately wheeled into a hospital's emergency room because your body and its stomach cannot handle that kind of alkalinity.

If you suffer from frequent heartburn or acid reflux disease (gastroesophageal reflux disease), drinking this 9.5 pH water will NOT help the least bit. Your stomach acids need to constantly stay very acidic to maintain proper digestive functions, and the slight alkalinity of this water has no effect on the stomach acid that is entering into the esophagus of people with GERD.

Trying to market bottled water based on its "silky smooth taste" and "hydrates better" claims, its over-hyped "9.5 pH" attribute, and all its overexaggerated health claims constitute marketing trickery to fool consumers by utilizing the "if you believe that it can really work, it does indeed work" placebo effect. The placebo effect is a very powerful phenomenon that can frequently affect everything from clinical trials of pharmacological substances to faith healing. Basically, if you have a hopeful mindset of expectancy and conditioning that makes you think that something will work, then that perception can become 100% of your subjectively observed reality and outcome of the treatment. For examples of other placebo effects: If I told you that you can get stronger bones by breathing milk through your nose instead of drinking it through your mouth, or if I told you that you could be revitalized by just holding broccoli in your hands for two minutes every day instead of eating it, if you really believe or hope that to be true, then after breathing the milk into your nose or holding the broccoli in your hands, you may think to yourself, "Gee, I feel better, stronger, and more energetic after having done that!" Likewise, if you really HOPE to feel more hydrated after drinking this Essentia water, you may actually feel more refreshed simply because you BELIEVE that this water is hydrating you better than drinking other sources of water, when in fact the tap water from your faucet could have hydrated you just as refreshingly.

The Amazon 'Product Description' for this Essentia water falsely states even more overexaggerated health claims: "Essentia Water provides unmatched hydration, health benefits and smooth taste. Its superior hydrating qualities come from a special electrolyte formula and optimal pH level of 9.5, which gives your body more of what it needs to thrive. Drinking Essentia Water boosts the antioxidant properties of your immune system and helps bring your body back into balance. It also helps maintain normal blood pressure, restful sleep, heart health, muscle strength and more." Again... these claims are wayyyyy overexaggerated! Compared to your overall health, fitness level, and daily food diet, the trace minerals found in this water and this water's slight alkalinity really make no difference in boosting the "antioxidant properties of your immune system", nor does it "help maintain normal blood pressure, restful sleep, heart health, muscle strength and more". With this litany of false health claims, I am surprised that they also did not pile on claims that this water "increases libido, cures erectile dysfunction, grows hair on balding men, makes women sexy, and fixes the national debt" too.

Fuzzy Wuzzy's Conclusion: I must admit that I personally do not regularly purchase bottled water for home use, even though I will sometimes buy it while I am away from home, both because I think that filtering my tap water through a Brita filter produces water that is just as tasty and healthy as bottled water and, just as important, I think that too much plastic (and petroleum) is used in the production of bottled water and most plastic bottles end up in our overflowing landfills. Roughly 1.5 million tons of plastic is used to make water bottles. And this does not include the energy used to make the plastic and the oil used for transporting all the bottled water. It takes about two liters of water used during the production process just to make one liter of bottled water. So behind each 1-liter bottle of water is another two liters of water that was expended during its production. Essentia states that their plastic bottles are "safe and recyclable", but we can all reduce our carbon footprint if we drink less bottled water. In many developing countries, tap water is either unavailable or unsafe, making bottled water a better option if residents cannot boil their tap water before consumption. But here in the U.S., some bottled water manufacturers, including this Essentia bottled water, use fear-based advertising claims that their water is healthier than tap water, which it really is not. The concerns for the taste of tap water are valid, however, especially with tap water that has residual smell or taste of chlorination. But residual smells and taste issues can be easily and economically remedied by using various activated charcoal/carbon filters at home to treat your tap water. Anyway, the tap water that comes out of my faucet at home is also slightly alkaline like this Essentia water, and many municipal water supplies are

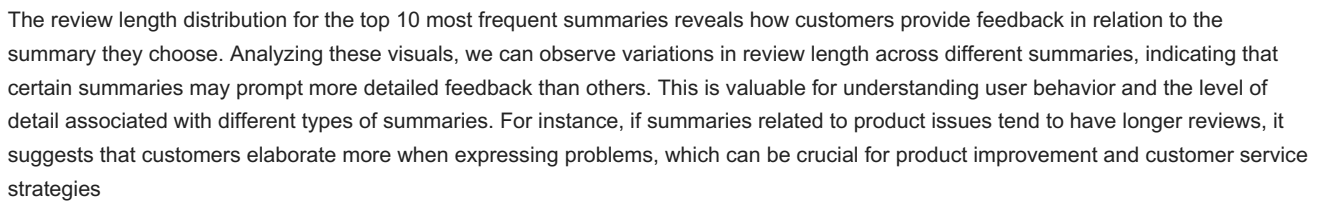
```
In [50]: import pandas as pd
import matplotlib.pyplot as plt

# Get the top 10 most frequent summaries
top_summaries = reviews['Summary'].value_counts().nlargest(10).index

# Filter dataset to include only those summaries
filtered_reviews = reviews[reviews['Summary'].isin(top_summaries)]

# Plot histograms grouped by top summaries
filtered_reviews.hist(column='length', by='Summary', bins=60, figsize=(12, 8), sharex=True, sharey=True)

plt.suptitle("Review Length Distribution for Top 10 Most Frequent Summaries", fontsize=15)
plt.show()
```



```
In [57]: #!nvidia-smi
```

```
In [58]: def text_process(mess):
        """
        1. Remove punctuation
        2. Remove stop words
        3. Return a list of cleaned texts.
        """
        clean_punctuation = [char for char in mess if char not in string.punctuation]
        clean_punctuation = ''.join(clean_punctuation)
```

```
return [word for word in clean_punctuation.split() if word.lower() not in stopwords.words('english')]
```

```
In [59]: reviews['Text'].head(20).apply(text_process)
```

```
Out[59]: 0    [bought, several, Vitality, canned, dog, food,...
1    [Product, arrived, labeled, Jumbo, Salted, Pea...
2    [confection, around, centuries, light, pillowy...
3    [looking, secret, ingredient, Robitussin, beli...
4    [Great, taffy, great, price, wide, assortment,...
5    [got, wild, hair, taffy, ordered, five, pound,...
6    [saltwater, taffy, great, flavors, soft, chewy...
7    [taffy, good, soft, chewy, flavors, amazing, w...
8    [Right, Im, mostly, sprouting, cats, eat, gras...
9    [healthy, dog, food, Good, digestion, Also, go...
10   [dont, know, cactus, tequila, unique, combinat...
11   [One, boys, needed, lose, weight, didnt, put, ...
12   [cats, happily, eating, Felidae, Platinum, two...
13   [good, flavor, came, securely, packed, fresh, ...
14   [Strawberry, Twizzlers, guilty, pleasure, yumm...
15   [daughter, loves, twizzlers, shipment, six, po...
16   [love, eating, good, watching, TV, looking, mo...
17   [satisfied, Twizzler, purchase, shared, others...
18   [Twizzlers, Strawberry, childhood, favorite, c...
19   [Candy, delivered, fast, purchased, reasonable...
Name: Text, dtype: object
```

```
In [60]: reviews[reviews['length']==21409]['Text'].iloc[0][:10]
```

```
Out[60]: 'Fuzzy Wuzz'
```

```
In [61]: reviews.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 568454 entries, 0 to 568453
Data columns (total 11 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Id                    568454 non-null  int64
1   ProductId            568454 non-null  object
2   UserId               568454 non-null  object
3   ProfileName          568454 non-null  object
4   HelpfulnessNumerator  568454 non-null  int64
5   HelpfulnessDenominator 568454 non-null  int64
6   Score                568454 non-null  int64
7   Time                 568454 non-null  int64
8   Summary              568454 non-null  object
9   Text                 568454 non-null  object
10  length               568454 non-null  int64
dtypes: int64(6), object(5)
memory usage: 47.7+ MB
```

```
In [62]: pip install textblob
```

```
Requirement already satisfied: textblob in c:\users\collins pc\anaconda3\envs\collonel\lib\site-packages (0.18.0.post0)Note: you may need to restart the kernel to use updated packages.
```

```
Requirement already satisfied: nltk>=3.8 in c:\users\collins pc\anaconda3\envs\collonel\lib\site-packages (from textblob) (3.8.1)
```

```
Requirement already satisfied: click in c:\users\collins pc\anaconda3\envs\collonel\lib\site-packages (from nltk>=3.8->textblob) (8.1.7)
```

```
Requirement already satisfied: joblib in c:\users\collins pc\anaconda3\envs\collonel\lib\site-packages (from nltk>=3.8->textblob) (1.4.0)
```

```
Requirement already satisfied: regex>=2021.8.3 in c:\users\collins pc\anaconda3\envs\collonel\lib\site-packages (from nltk>=3.8->textblob) (2024.4.16)
```

```
Requirement already satisfied: tqdm in c:\users\collins pc\anaconda3\envs\collonel\lib\site-packages (from nltk>=3.8->textblob) (4.66.2)
```

```
Requirement already satisfied: colorama in c:\users\collins pc\anaconda3\envs\collonel\lib\site-packages (from click->nltk>=3.8->textblob) (0.4.6)
```

```
In [63]: !pip install swifter --index-url https://pypi.org/simple/
```

Looking in indexes: <https://pypi.org/simple/>
Requirement already satisfied: swifter in c:\users\collins pc\anaconda3\envs\collonel\lib\site-packages (1.4.0)
Requirement already satisfied: pandas>=1.0.0 in c:\users\collins pc\anaconda3\envs\collonel\lib\site-packages (from swifter) (2.2.2)
Requirement already satisfied: psutil>=5.6.6 in c:\users\collins pc\anaconda3\envs\collonel\lib\site-packages (from swifter) (5.9.0)
Requirement already satisfied: dask>=2.10.0 in c:\users\collins pc\anaconda3\envs\collonel\lib\site-packages (from dask[dataframe]>=2.10.0->swifter) (2025.2.0)
Requirement already satisfied: tqdm>=4.33.0 in c:\users\collins pc\anaconda3\envs\collonel\lib\site-packages (from swifter) (4.66.2)
Requirement already satisfied: click>=8.1 in c:\users\collins pc\anaconda3\envs\collonel\lib\site-packages (from dask>=2.10.0->dask[dataframe]>=2.10.0->swifter) (8.1.7)
Requirement already satisfied: cloudpickle>=3.0.0 in c:\users\collins pc\anaconda3\envs\collonel\lib\site-packages (from dask>=2.10.0->dask[dataframe]>=2.10.0->swifter) (3.1.1)
Requirement already satisfied: fsspec>=2021.09.0 in c:\users\collins pc\anaconda3\envs\collonel\lib\site-packages (from dask>=2.10.0->dask[dataframe]>=2.10.0->swifter) (2025.2.0)
Requirement already satisfied: packaging>=20.0 in c:\users\collins pc\anaconda3\envs\collonel\lib\site-packages (from dask>=2.10.0->dask[dataframe]>=2.10.0->swifter) (23.2)
Requirement already satisfied: partd>=1.4.0 in c:\users\collins pc\anaconda3\envs\collonel\lib\site-packages (from dask>=2.10.0->dask[dataframe]>=2.10.0->swifter) (1.4.2)
Requirement already satisfied: pyyaml>=5.3.1 in c:\users\collins pc\anaconda3\envs\collonel\lib\site-packages (from dask>=2.10.0->dask[dataframe]>=2.10.0->swifter) (6.0.1)
Requirement already satisfied: toolz>=0.10.0 in c:\users\collins pc\anaconda3\envs\collonel\lib\site-packages (from dask>=2.10.0->dask[dataframe]>=2.10.0->swifter) (1.0.0)
Requirement already satisfied: importlib_metadata>=4.13.0 in c:\users\collins pc\anaconda3\envs\collonel\lib\site-packages (from dask>=2.10.0->dask[dataframe]>=2.10.0->swifter) (8.5.0)
Requirement already satisfied: pyarrow>=14.0.1 in c:\users\collins pc\anaconda3\envs\collonel\lib\site-packages (from dask[dataframe]>=2.10.0->swifter) (19.0.1)
Requirement already satisfied: numpy>=1.23.2 in c:\users\collins pc\anaconda3\envs\collonel\lib\site-packages (from pandas>=1.0.0->swifter) (1.26.4)
Requirement already satisfied: python-dateutil>=2.8.2 in c:\users\collins pc\anaconda3\envs\collonel\lib\site-packages (from pandas>=1.0.0->swifter) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in c:\users\collins pc\anaconda3\envs\collonel\lib\site-packages (from pandas>=1.0.0->swifter) (2023.3.post1)
Requirement already satisfied: tzdata>=2022.7 in c:\users\collins pc\anaconda3\envs\collonel\lib\site-packages (from pandas>=1.0.0->swifter) (2024.1)
Requirement already satisfied: colorama in c:\users\collins pc\anaconda3\envs\collonel\lib\site-packages (from tqdm>=4.33.0->swifter) (0.4.6)
Requirement already satisfied: zipp>=3.20 in c:\users\collins pc\anaconda3\envs\collonel\lib\site-packages (from importlib_metadata>=4.13.0->dask>=2.10.0->dask[dataframe]>=2.10.0->swifter) (3.21.0)
Requirement already satisfied: locket in c:\users\collins pc\anaconda3\envs\collonel\lib\site-packages (from partd>=1.4.0->dask>=2.10.0->dask[dataframe]>=2.10.0->swifter) (1.0.0)
Requirement already satisfied: six>=1.5 in c:\users\collins pc\anaconda3\envs\collonel\lib\site-packages (from python-dateutil>=2.8.2->pandas>=1.0.0->swifter) (1.16.0)

```
In [68]: # Jupyter Notebook
import pandas as pd
import importlib
import algorithms
importlib.reload(algorithms)
from algorithms import analyze_sentiment #Algorithm that computes the sentiment
# analysis from a large Copora of texts such as the reviews texts with 568,454 reviews

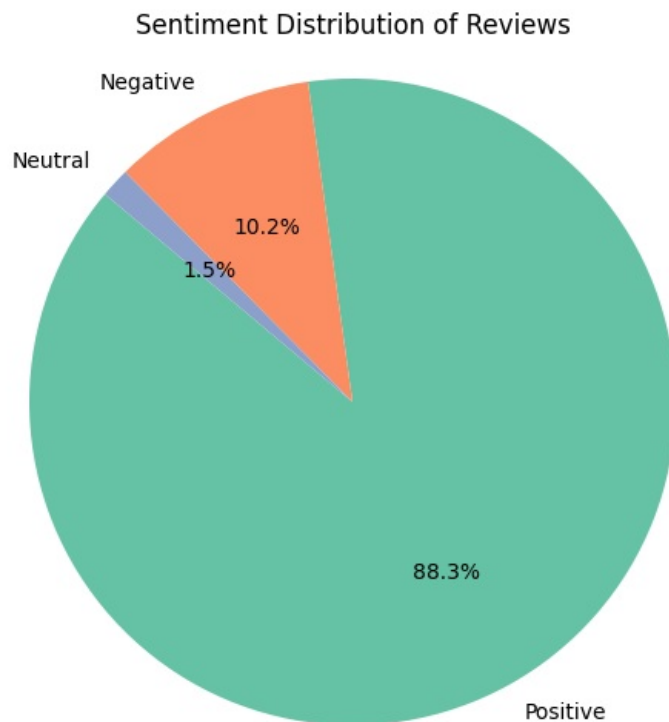
# Pick reviews from a previous analysis
reviews = reviews['Text'].to_list() # Convert to list

positive, negative, neutral, total = analyze_sentiment(reviews) #Implimenting the Algorithm

# Dispalying the results of the sentiment analysis

print(f"\nSentiment Analysis Results (from notebook):")
print(f"Total Reviews: {total}")
print(f"Positive: {positive}, Negative: {negative}, Neutral: {neutral}")
```

Out of 568454 reviews:
502016 were positive
58051 were negative
8387 were neutral



Sentiment Analysis Results (from notebook):
 Total Reviews: 568454
 Positive: 502016, Negative: 58051, Neutral: 8387

The sentiment analysis of **568,454 customer reviews** reveals a strong **positive sentiment trend**, with **502,016 reviews** classified as positive compared to **58,051 negative** and **8,387 neutral reviews**. This indicates that the majority of customers are satisfied, which can be leveraged to highlight popular products and positive feedback in **marketing strategies**. However, the significant number of **negative reviews** also points to **areas needing improvement**, as addressing these concerns could enhance **overall customer satisfaction** and potentially **drive increased sales**. Understanding the specific issues raised in negative reviews can inform targeted product development or customer service initiatives.

```
In [70]: import matplotlib.pyplot as plt
from collections import Counter
import re
import nltk
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize

# Download necessary NLTK data
nltk.download('stopwords')
nltk.download('punkt')

# Preprocess text data
stop_words = set(stopwords.words('english'))

def preprocess_text(text):
    if not text: # Handle empty strings
        return []
    text = re.sub(r'^a-zA-Z\s', '', text) # Remove non-alphabetic chars
    words = word_tokenize(text.lower()) # Tokenize and lowercase
    words = [word for word in words if word not in stop_words] # Remove stopwords
    return words

# Process reviews list
all_words = []
```

```

for review in reviews:
    all_words.extend(preprocess_text(review))

# Get word frequencies
word_counts = Counter(all_words)
most_common_words = word_counts.most_common(10)
words, frequencies = zip(*most_common_words)

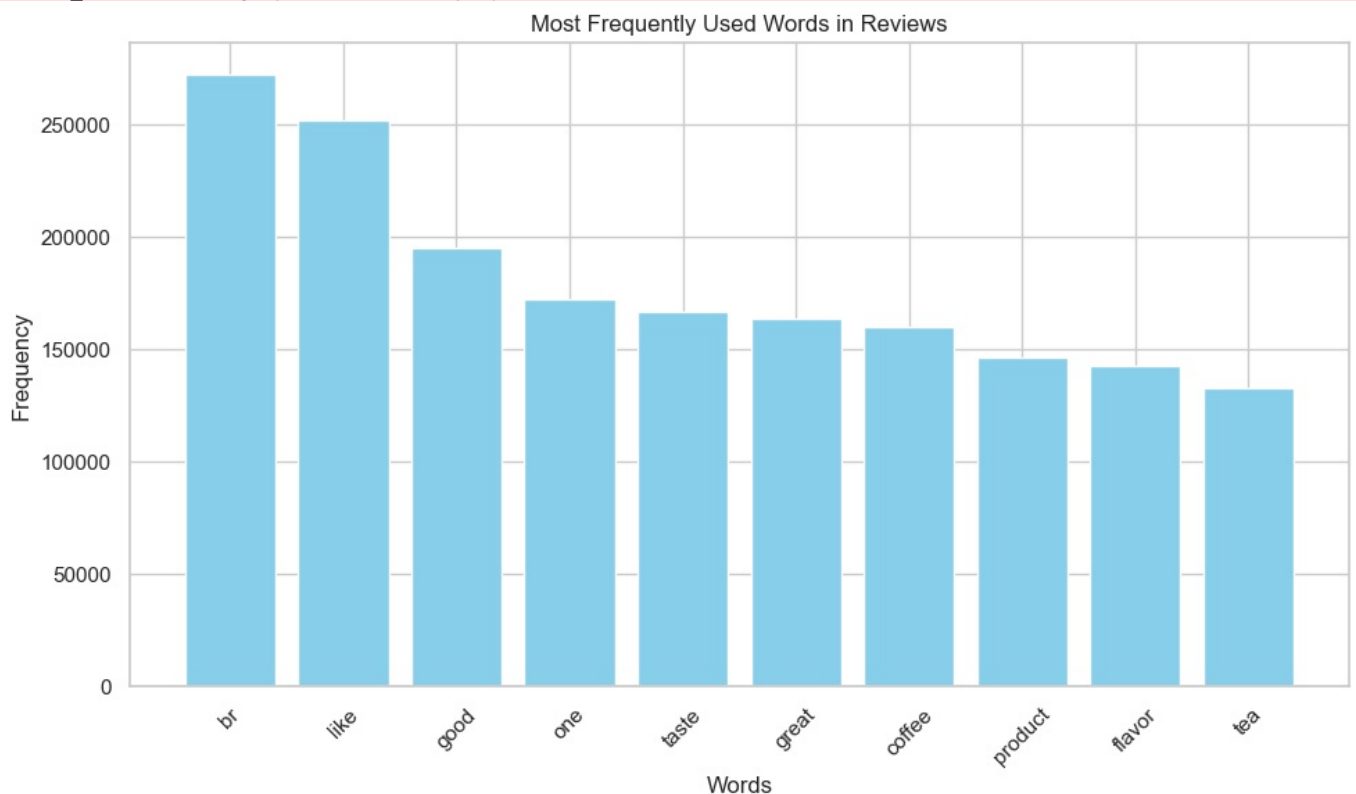
# Plotting
plt.figure(figsize=(10, 6))
plt.bar(words, frequencies, color='skyblue')
plt.xlabel('Words')
plt.ylabel('Frequency')
plt.title('Most Frequently Used Words in Reviews')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()

```

```

[nltk_data] Downloading package stopwords to C:\Users\Collins
[nltk_data]   PC\AppData\Roaming\nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
[nltk_data] Downloading package punkt to C:\Users\Collins
[nltk_data]   PC\AppData\Roaming\nltk_data...
[nltk_data]   Package punkt is already up-to-date!

```



The analysis of most common word frequencies reveals **key themes in customer reviews**, with prominent terms including 'like,' 'good,' 'taste,' 'great,' and 'coffee,' indicating a **strong focus on product quality and sensory experience**. The prevalence of 'like' and 'good' suggests frequent expressions of **general approval**, while 'taste,' 'great,' and 'flavor' highlight the **importance of product palatability and enjoyment**. Furthermore, specific terms like 'coffee' and 'tea' point to **product categories that drive much of the discussion**. This insight allows businesses to understand **core customer interests**, refine product development to enhance taste and quality, and tailor marketing messages to emphasize positive sensory attributes and popular product categories.

```
In [33]: reviews[:10]
```



```
Out[33]: ['I have bought several of the Vitality canned dog food products and have found them all to be of good quality. The product looks more like a stew than a processed meat and it smells better. My Labrador is finicky and she appreciates this product better than most.',
          'Product arrived labeled as Jumbo Salted Peanuts...the peanuts were actually small sized unsalted. Not sure if this was an error or if the vendor intended to represent the product as "Jumbo".',
          'This is a confection that has been around a few centuries. It is a light, pillowy citrus gelatin with nuts - in this case Filberts. And it is cut into tiny squares and then liberally coated with powdered sugar. And it is a tiny mouthful of heaven. Not too chewy, and very flavorful. I highly recommend this yummy treat. If you are familiar with the story of C.S. Lewis\' "The Lion, The Witch, and The Wardrobe" - this is the treat that seduces Edmund into selling out his Brother and Sisters to the Witch.',
          'If you are looking for the secret ingredient in Robitussin I believe I have found it. I got this in addition to the Root Beer Extract I ordered (which was good) and made some cherry soda. The flavor is very medicinal.',
          'Great taffy at a great price. There was a wide assortment of yummy taffy. Delivery was very quick. If you are a taffy lover, this is a deal.',
          'I got a wild hair for taffy and ordered this five pound bag. The taffy was all very enjoyable with many flavors: watermelon, root beer, melon, peppermint, grape, etc. My only complaint is there was a bit too much red/black licorice-flavored pieces (just not my particular favorites). Between me, my kids, and my husband, this lasted only two weeks! I would recommend this brand of taffy -- it was a delightful treat.',
          "This saltwater taffy had great flavors and was very soft and chewy. Each candy was individually wrapped well. None of the candies were stuck together, which did happen in the expensive version, Fralinger's. Would highly recommend this candy! I served it at a beach-themed party and everyone loved it!",
          'This taffy is so good. It is very soft and chewy. The flavors are amazing. I would definitely recommend you buying it. Very satisfying!!',
          "Right now I'm mostly just sprouting this so my cats can eat the grass. They love it. I rotate it around with Wheatgrass and Rye too",
          'This is a very healthy dog food. Good for their digestion. Also good for small puppies. My dog eats her required amount at every feeding.']
```

```
In [51]: print(len(reviews))
```

568454

```
In [72]: import sqlite3
import pandas as pd

# Connect to SQLite database
conn = sqlite3.connect("database.sqlite")

# Read entire Reviews table into a Pandas DataFrame
df = pd.read_sql("SELECT * FROM Reviews", conn)

# Close the connection
conn.close()

# Display the first few rows
print(df.head())
```

	Id	ProductId	UserId	ProfileName	\
0	1	B001E4KFG0	A3SGXH7AUHU8GW	delmartian	
1	2	B00813GRG4	A1D87F6ZCVE5NK	dll pa	
2	3	B000LQOCH0	ABXLMWJIXXAIN	Natalia Corres	"Natalia Corres"
3	4	B000UA0QIQ	A395B0RC6FGVXV	Karl	
4	5	B006K2ZZ7K	A1UQRSCLF8GW1T	Michael D. Bigham	"M. Wassir"

	HelpfulnessNumerator	HelpfulnessDenominator	Score	Time	\
0	1	1	5	1303862400	
1	0	0	1	1346976000	
2	1	1	4	1219017600	
3	3	3	2	1307923200	
4	0	0	5	1350777600	

	Summary	Text
0	Good Quality Dog Food	I have bought several of the Vitality canned d...
1	Not as Advertised	Product arrived labeled as Jumbo Salted Peanut...
2	"Delight" says it all	This is a confection that has been around a fe...
3	Cough Medicine	If you are looking for the secret ingredient i...
4	Great taffy	Great taffy at a great price. There was a wid...

```
In [73]: import matplotlib.pyplot as plt
import seaborn as sns

# Set seaborn style
sns.set_style("whitegrid")

# Convert value counts to a DataFrame
score_counts = df['Score'].value_counts().sort_index().reset_index()
score_counts.columns = ['Score', 'Count']

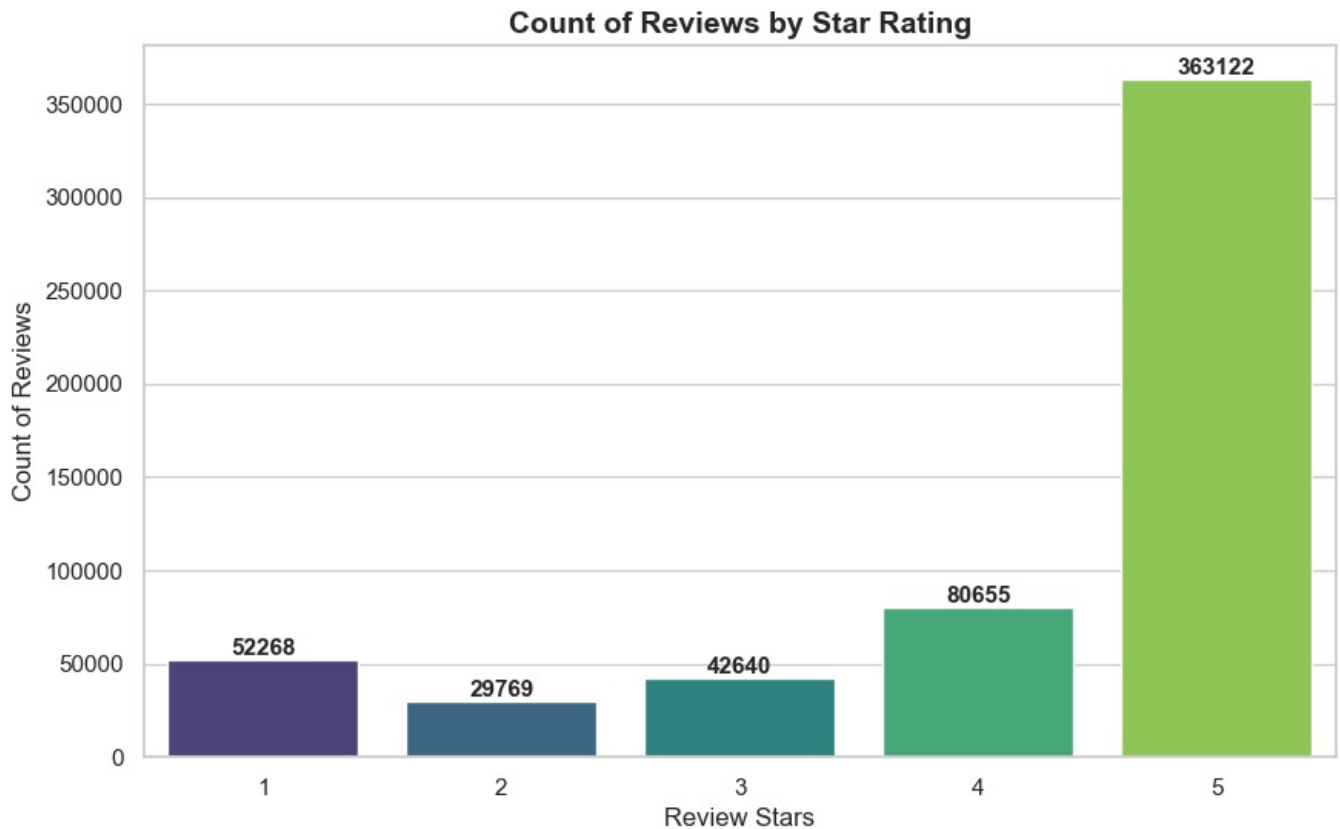
# Create figure
plt.figure(figsize=(10, 6))

# Plot using seaborn
ax = sns.barplot(data=score_counts, x='Score', y='Count', palette="viridis")
```

```
# Add labels
plt.title('Count of Reviews by Star Rating', fontsize=14, fontweight='bold')
plt.xlabel('Review Stars', fontsize=12)
plt.ylabel('Count of Reviews', fontsize=12)

# Show values on bars
for p in ax.patches:
    ax.annotate(f'{int(p.get_height())}',
                (p.get_x() + p.get_width() / 2, p.get_height()),
                ha='center', va='bottom', fontsize=11, fontweight='bold')

# Show plot
plt.show()
```



The distribution of review scores clearly indicates a strong **positive sentiment bias**, with the majority of **reviewers awarding the highest rating (5)**, significantly **outnumbering all other scores**. While there's a notable drop-off in counts for scores 4 through 1, the overall trend suggests that **most customers are generally satisfied**. This insight is crucial for businesses to acknowledge their strengths and maintain positive aspects of their products or services. However, it's **equally important to examine the reviews associated with the less frequent lower scores (1-3)** to identify **areas for potential improvement** and address customer concerns, even if they represent a smaller portion of the overall feedback.

```
In [73]: import swifter
df['Processed_Text'] = df['Text'].dropna().swifter.apply(text_process)
```

Pandas Apply: 0% | 0/568454 [00:00<?, ?it/s]

```
In [74]: df
```

	Id	ProductId	UserId	ProfileName	HelpfulnessNumerator	HelpfulnessDenominator	Score	Time
0	1	B001E4KFG0	A3SGXH7AUHU8GW	delmartian	1	1	5	130386240
1	2	B00813GRG4	A1D87F6ZCVE5NK	dll pa	0	0	1	134697600
2	3	B000LQOCH0	ABXLMWJIXXAIN	Natalia Corres "Natalia Corres"	1	1	4	121901760
3	4	B000UA0QIQ	A395BORC6FGVXV	Karl	3	3	2	130792320
4	5	B006K2ZZ7K	A1UQRSCLF8GW1T	Michael D. Bigham "M. Wassir"	0	0	5	135077760
...
568449	568450	B001EO7N10	A28KG5XORO54AY	Lettie D. Carter	0	0	5	129962880
568450	568451	B003S1WTCU	A3I8AFVPEE8KI5	R. Sawyer	0	0	2	133125120
568451	568452	B004I613EE	A121AA1GQV751Z	pkdsd "pk_007"	2	2	5	132978240
568452	568453	B004I613EE	A3IBEVCTXKNOH	Kathy A. Welch "katwel"	1	1	5	133159680
568453	568454	B001LR2CU2	A3LGQPJCZVL9UC	srfell17	0	0	5	133842240

Model Building

```
In [35]: import pandas as pd

#convert reviews into a dataframe
df_reviews = pd.DataFrame({'review': reviews})

# Display the first few rows
print(df_reviews.head())
```

```

                                review
0  I have bought several of the Vitality canned d...
1  Product arrived labeled as Jumbo Salted Peanut...
2  This is a confection that has been around a fe...
3  If you are looking for the secret ingredient i...
4  Great taffy at a great price. There was a wid...

```

```
In [36]: df_reviews.head()
```

```

Out[36]:
                                review
0  I have bought several of the Vitality canned d...
1  Product arrived labeled as Jumbo Salted Peanut...
2  This is a confection that has been around a fe...
3  If you are looking for the secret ingredient i...
4  Great taffy at a great price. There was a wid...

```

```
In [37]: # Add the 'summary' column from df to df_reviews
df_reviews['Summary'] = df['Summary']
```

```
In [38]: df_reviews
```

```

Out[38]:
                                review                                Summary
0  I have bought several of the Vitality canned d...  Good Quality Dog Food
1  Product arrived labeled as Jumbo Salted Peanut...  Not as Advertised
2  This is a confection that has been around a fe...  "Delight" says it all
3  If you are looking for the secret ingredient i...  Cough Medicine
4  Great taffy at a great price. There was a wid...  Great taffy
...
568449  Great for sesame chicken..this is a good if no...  Will not do without
568450  I'm disappointed with the flavor. The chocolat...  disappointed
568451  These stars are small, so you can give 10-15 o...  Perfect for our maltipoo
568452  These are the BEST treats for training and rew...  Favorite Training and reward treat
568453  I am very satisfied ,product is as advertised,...  Great Honey

```

568454 rows × 2 columns

```
In [39]: df_reviews['Score'] = df['Score']
```

```
In [40]: df_reviews
```

```

Out[40]:
                                review                                Summary  Score
0  I have bought several of the Vitality canned d...  Good Quality Dog Food      5
1  Product arrived labeled as Jumbo Salted Peanut...  Not as Advertised      1
2  This is a confection that has been around a fe...  "Delight" says it all      4
3  If you are looking for the secret ingredient i...  Cough Medicine          2
4  Great taffy at a great price. There was a wid...  Great taffy             5
...
568449  Great for sesame chicken..this is a good if no...  Will not do without      5
568450  I'm disappointed with the flavor. The chocolat...  disappointed              2
568451  These stars are small, so you can give 10-15 o...  Perfect for our maltipoo  5
568452  These are the BEST treats for training and rew...  Favorite Training and reward treat  5
568453  I am very satisfied ,product is as advertised,...  Great Honey              5

```

568454 rows × 3 columns

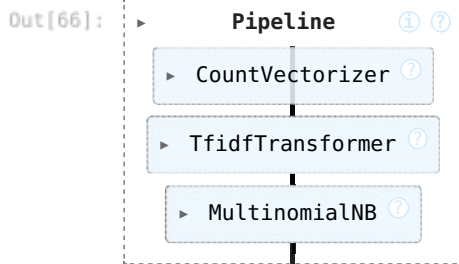
Multinomial Naive Bayes

```
In [63]: review_train, review_test, stars_train, stars_test = train_test_split(
    df_reviews['review'],
    df_reviews['Score'],
    test_size=0.3,
```

```
    random_state=42
)
```

```
In [65]: from sklearn.pipeline import Pipeline
pipeline = Pipeline([
    ('bow', CountVectorizer(analyzer=text_process)), # text to token counts
    ('tfidf', TfidfTransformer()),                 # normalize
    ('classifier', MultinomialNB())                 # classifier
])
```

```
In [66]: pipeline.fit(review_train, stars_train)
```



```
In [67]: predictions = pipeline.predict(review_test)
```

```
In [68]: print(classification_report(stars_test, predictions))
```

	precision	recall	f1-score	support
1	0.86	0.05	0.09	15513
2	0.88	0.00	0.00	8765
3	1.00	0.00	0.00	12850
4	0.87	0.01	0.02	24286
5	0.64	1.00	0.78	109123
accuracy			0.65	170537
macro avg	0.85	0.21	0.18	170537
weighted avg	0.74	0.65	0.51	170537

```
In [69]: predictions
```

```
Out[69]: array([5, 5, 5, ..., 5, 5, 5], dtype=int64)
```

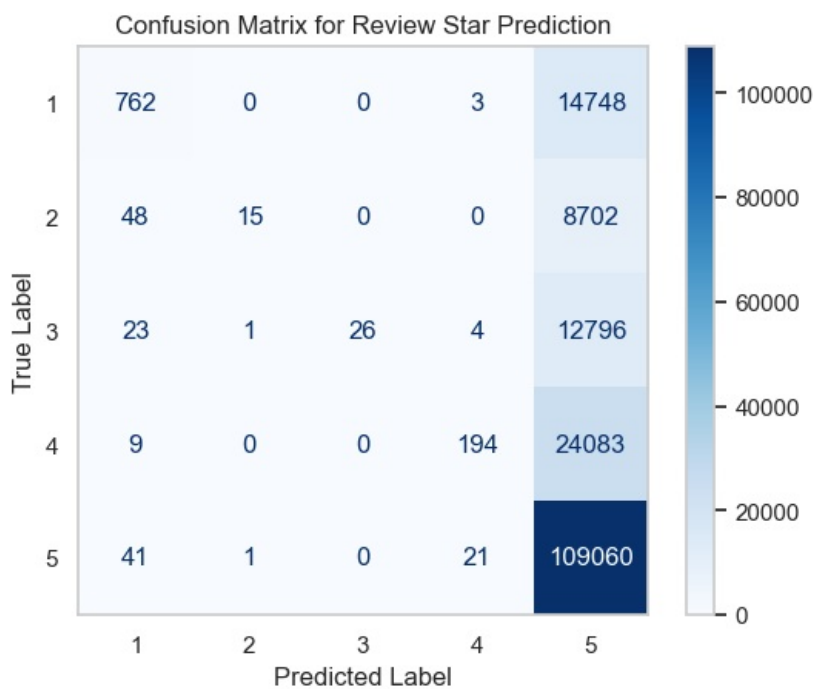
```
In [70]: import matplotlib.pyplot as plt
from sklearn.metrics import confusion_matrix, ConfusionMatrixDisplay

# Step 1: Compute the confusion matrix
cm = confusion_matrix(stars_test, predictions)

# Step 2: Create a display object
disp = ConfusionMatrixDisplay(confusion_matrix=cm, display_labels=[1, 2, 3, 4, 5])

# Step 3: Plot it
plt.figure(figsize=(8, 6))
disp.plot(cmap=plt.cm.Blues, values_format='d') # 'd' for integer values
plt.title('Confusion Matrix for Review Star Prediction')
plt.xlabel('Predicted Label')
plt.ylabel('True Label')
plt.grid(False)
plt.show()
```

<Figure size 800x600 with 0 Axes>



Model Performance

The Multinomial Naive Bayes model was used to classify customer reviews. The classification report reveals the following performance metrics:

Accuracy:

The overall accuracy of the model is **65%**.

Precision:

Precision varies significantly across classes, ranging from **64% to 100%**. The model shows **high precision for classes 1, 2, 3, and 4 (86%, 88%, 100%, and 87% respectively)**, but lower precision for class 5 (64%). This indicates that when the model predicts a **review belongs to classes 1-4**, it is **generally correct**, but it is **less reliable when predicting class 5**.

Recall:

Recall also varies widely, from **0% to 100%**. The model has very **low recall for classes 1, 2, 3, and 4 (5%, 0%, 0%, and 1% respectively)**, indicating that it **fails to correctly identify most reviews belonging to these classes**. In contrast, the recall for **class 5 is 100%**, meaning the **model correctly identifies all reviews of this class**.

F1-Score:

The F1-score, which balances precision and recall, is low for classes 1, 2, 3, and 4 (**0.09, 0.00, 0.00, and 0.02 respectively**) and high for **class 5 (0.78)**. This confirms the model's **poor performance in accurately classifying classes 1-4**.

Pros of Using Multinomial Naive Bayes

Simplicity and Speed:

Multinomial NB is computationally **efficient and easy to implement**. And given the large size of the dataset it **performed much faster than the Random Forest Classifier**.

Effective with Text Data:

It is **well-suited** for **text classification tasks**, especially with high-dimensional data.

Cons of Using Multinomial Naive Bayes

Poor Performance on Imbalanced Data:

The classification report indicates that the model **struggles with imbalanced datasets**. The high recall for class 5 and low recall for other classes suggest that the **model is biased towards the majority class**.

Zero Frequency Problem:

Naive Bayes can assign **zero probability to words not seen in the training data**, although this can be mitigated with **smoothing techniques**.

Independence Assumption:

Naive Bayes **assumes** that **features are independent**, which is often not the case in text data.

Conclusion

The Multinomial Naive Bayes model's **performance is heavily influenced by the class imbalance** in the dataset. While it **correctly classifies reviews** in the **majority class (class 5)**, it performs **poorly in identifying reviews from other classes**. To improve performance, it may be necessary to address the **class imbalance** through techniques like **oversampling**, **undersampling**, or using **different algorithms** that are more **robust to imbalanced data**.

Random Forest Classifier

```
In [45]: from sklearn.ensemble import RandomForestClassifier
from sklearn.feature_extraction.text import CountVectorizer, TfidfTransformer
from sklearn.pipeline import Pipeline
from sklearn.model_selection import train_test_split
from sklearn.metrics import classification_report, confusion_matrix
import matplotlib.pyplot as plt
```

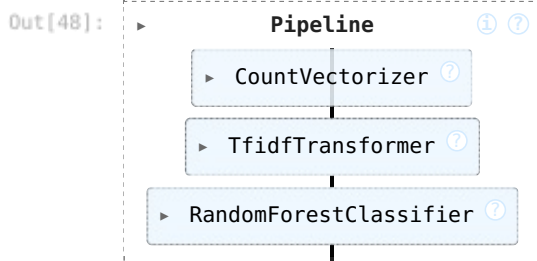
```
In [46]: # Features and target
X = df_reviews['review']
y = df_reviews['Score']

# Train-test split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)
```

```
In [47]: from sklearn.ensemble import RandomForestClassifier

pipeline_rf = Pipeline([
    ('bow', CountVectorizer(analyzer=text_process)),           # Bag of Words
    ('tfidf', TfidfTransformer()),                          # TF-IDF
    ('classifier', RandomForestClassifier(n_estimators=100, random_state=42)) # Random Forest
])
```

```
In [48]: pipeline_rf.fit(X_train, y_train)
```



```
In [49]: predictions_rf = pipeline_rf.predict(X_test)
```

```
In [50]: print(classification_report(y_test, predictions_rf))
```

	precision	recall	f1-score	support
1	0.95	0.49	0.64	15513
2	1.00	0.39	0.56	8765
3	0.98	0.38	0.54	12850
4	0.99	0.38	0.55	24286
5	0.75	1.00	0.86	109123
accuracy			0.79	170537
macro avg	0.93	0.53	0.63	170537
weighted avg	0.83	0.79	0.76	170537

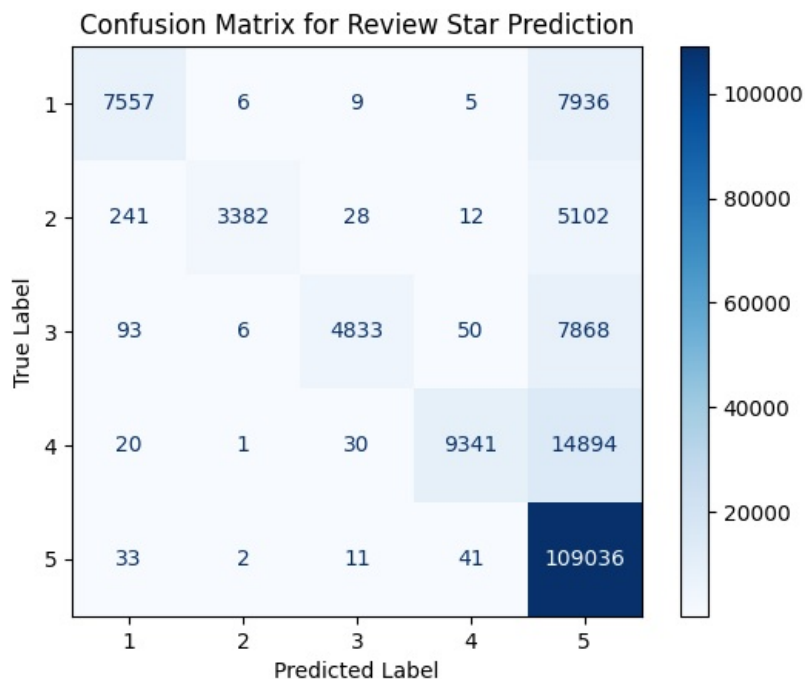
```
In [51]: import matplotlib.pyplot as plt
from sklearn.metrics import confusion_matrix, ConfusionMatrixDisplay

# Step 1: Compute the confusion matrix
cm = confusion_matrix(y_test, predictions_rf)

# Step 2: Create a display object
disp = ConfusionMatrixDisplay(confusion_matrix=cm, display_labels=[1, 2, 3, 4, 5])

# Step 3: Plot it
plt.figure(figsize=(8, 6))
disp.plot(cmap=plt.cm.Blues, values_format='d') # 'd' for integer values
plt.title('Confusion Matrix for Review Star Prediction')
plt.xlabel('Predicted Label')
plt.ylabel('True Label')
plt.grid(False)
plt.show()
```

<Figure size 800x600 with 0 Axes>



Model Performance

The Random Forest model was used to classify customer reviews, and the classification report shows the following:

Accuracy:

The overall accuracy of the model is **79%**.

Precision:

Precision is **high across classes 1-4**, ranging from **95% to 100%**. Precision for class 5 is 75%. This indicates that when the model predicts a review **belongs to classes 1-4, it is highly reliable**. The precision for **class 5 is still reasonably good**, suggesting that the model is fairly accurate in predicting reviews for this class as well.

Recall:

Recall varies across classes, from 38% to 100%. The model demonstrates lower recall for classes 1-4 (38% to 49%), indicating that it misses a significant portion of reviews from these classes. However, the recall for class 5 is **100%, meaning the model correctly identifies all reviews** belonging to this class.

F1-Score:

The F1-score, balancing precision and recall, ranges from **0.54 to 0.86**. Classes 1-4 have F1-scores between 0.54 and 0.64, while class 5 has a higher F1-score of 0.86.

Comparison to Naive Bayes Model

Accuracy:

The **Random Forest** model (**79%**) **outperforms** the Naive Bayes model (**65%**) in terms of overall accuracy.

Precision:

Both models exhibit **high precision** for classes 1-4. However, **Random Forest generally maintains higher precision across these classes**. For class 5, Random Forest has a higher precision (75%) compared to Naive Bayes (64%).

Recall:

Naive Bayes shows a more extreme contrast in recall, with very low recall for classes 1-4 and 100% recall for class 5. Random Forest also has 100% recall for class 5 but has considerably better recall for classes 1-4 compared to Naive Bayes.

F1-Score:

Random Forest consistently achieves higher F1-scores across all classes compared to Naive Bayes, indicating a better balance between precision and recall.

Conclusion

The Random Forest model demonstrates a **significant improvement** over the **Naive Bayes model** in this classification project. It provides higher **accuracy**, **better precision for class 5**, and **improved recall and F1-scores** across classes **1-4**. While both models **struggle with class imbalance** (as indicated by the high recall for class 5), Random Forest **handles it more effectively**, delivering a **more balanced and accurate classification**.

In []:

In []:

In []:

In []:

In []:

In []:

In []: