

Amazon India Reviews Project

Brief Description of the Project: Performed an end-to-end analysis of Amazon India product reviews using Python. Conducted exploratory data analysis(EDA) and sentiment analysis using Natural Language Processing(NLP) tools on Amazon India product reviews. Applied NLP techniques including tokenization, stopword removal, and sentiment scoring with NLTK VADER and Hugging Face BERT based pretrained RoBERTa transformer. Created visualizations such as word clouds and statistical plots with Matplotlib and Seaborn. Demonstrated skills in data cleaning, text processing, sentiment analysis, and data visualization using libraries including pandas, numpy, NLTK, pytorch, transformers, Matplotlib and Seaborn to extract business insights and create impactful visualizations.

Importing necessary libraries:

```
In [4]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

import nltk
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
from nltk.sentiment import SentimentIntensityAnalyzer
from wordcloud import WordCloud

import torch
from transformers import AutoTokenizer, AutoModelForSequenceClassification

import warnings
warnings.filterwarnings('ignore')
```

Reading the dataset:

```
In [6]: df = pd.read_csv(r"D:\Data Science\Projects\2. Amazon India Reviews Project\amazon.csv")
```

```
In [7]: df.head()
```

	product_id	product_name	category	discounted_price	actual_price	discount_percentage	rating	ratio
0	B07JW9H4J1	Wayona Nylon Braided USB to Lightning Fast Charge	Computers&Accessories Accessories&Peripherals ...	₹399	₹1,099	64%	4.2	
1	B098NS6PVG	Ambrane Unbreakable 60W / 3A Fast Charging 1.5m	Computers&Accessories Accessories&Peripherals ...	₹199	₹349	43%	4.0	
2	B096MSW6CT	Sounce Fast Phone Charging Cable & Data Sync U...	Computers&Accessories Accessories&Peripherals ...	₹199	₹1,899	90%	3.9	
3	B08HDJ86NZ	boAt Deuce USB 300 2 in 1 Type-C & Micro USB S...	Computers&Accessories Accessories&Peripherals ...	₹329	₹699	53%	4.2	
4	B08CF3B7N1	Portronics Konnect L 1.2M Fast Charging 3A 8 P...	Computers&Accessories Accessories&Peripherals ...	₹154	₹399	61%	4.2	

```
In [8]: df.columns
```

```
Out[8]: Index(['product_id', 'product_name', 'category', 'discounted_price',
       'actual_price', 'discount_percentage', 'rating', 'rating_count',
       'about_product', 'user_id', 'user_name', 'review_id', 'review_title',
       'review_content', 'img_link', 'product_link'],
      dtype='object')
```

```
In [9]: df.shape
```

```
Out[9]: (1465, 16)
```

```
In [10]: df.size
```

```
Out[10]: 23440
```

```
In [11]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1465 entries, 0 to 1464
Data columns (total 16 columns):
 #   Column           Non-Null Count  Dtype  
---  --  
0   product_id      1465 non-null    object  
1   product_name    1465 non-null    object  
2   category        1465 non-null    object  
3   discounted_price 1465 non-null    object  
4   actual_price    1465 non-null    object  
5   discount_percentage 1465 non-null    object  
6   rating          1465 non-null    object  
7   rating_count    1463 non-null    object  
8   about_product   1465 non-null    object  
9   user_id         1465 non-null    object  
10  user_name       1465 non-null    object  
11  review_id       1465 non-null    object  
12  review_title   1465 non-null    object  
13  review_content  1465 non-null    object  
14  img_link        1465 non-null    object  
15  product_link   1465 non-null    object  
dtypes: object(16)
memory usage: 183.3+ KB
```

Observations:

There are 1465 rows and 16 columns in the dataset.

The data type of all columns is object.

The columns in the datasets are: 'product_id', 'product_name', 'category', 'discounted_price', 'actual_price', 'discount_percentage', 'rating', 'rating_count', 'about_product', 'user_id', 'user_name', 'review_id', 'review_title', 'review_content', 'img_link', 'product_link'

Changing data types of numeric columns:

```
In [14]: def convert_to_float(df, columns):

    for col in columns:
        df[col] = (
            df[col]
            .astype(str) # We make sure each value in the column is treated as a string
            .str.replace(r"[^\d\.\.]", "", regex=True) # This matches anything that's not a digit or period and replaces it with a empty string
            .replace("", "0") # Replace empty strings with zero
            .astype(float) # we now convert the cleaned string to a float data type
        )
    return df

numeric_columns = ['discounted_price', 'actual_price', 'discount_percentage', 'rating', 'rating_count']
df = convert_to_float(df, numeric_columns)

df.sample(5)
```

Out[14]:

	product_id	product_name	category	discounted_price	actual_price	discount_percentage	rat
242	B0758F7KK7	Caprigo Heavy Duty TV Wall Mount Bracket for 1...	Electronics HomeTheater,TV&Video Accessories T...	399.0	999.0	60.0	
1225	B09SZ5TWHW	Swiss Military VC03 Wireless Car Vacuum Cleane...	Home&Kitchen Kitchen&HomeAppliances Vacuum,Cle...	1547.0	2890.0	46.0	
1128	B00E9G8KOY	HUL Pureit Germkill kit for Classic 23 L water...	Home&Kitchen Kitchen&HomeAppliances WaterPurif...	600.0	600.0	0.0	
0	B07JW9H4J1	Wayona Nylon Braided USB to Lightning Fast Cha...	Computers&Accessories Accessories&Peripherals ...	399.0	1099.0	64.0	
1218	B0B9F9PT8R	Eopora PTC Ceramic Fast Heating Room Heater fo...	Home&Kitchen Heating,Cooling&AirQuality RoomHe...	1529.0	2999.0	49.0	

In [15]: `df.describe().round(2)`

Out[15]:

	discounted_price	actual_price	discount_percentage	rating	rating_count
count	1465.00	1465.00	1465.00	1465.00	1465.00
mean	3125.31	5444.99	47.69	4.09	18270.56
std	6944.30	10874.83	21.64	0.31	42730.00
min	39.00	39.00	0.00	0.00	0.00
25%	325.00	800.00	32.00	4.00	1173.00
50%	799.00	1650.00	50.00	4.10	5178.00
75%	1999.00	4295.00	63.00	4.30	17325.00
max	77990.00	139900.00	94.00	5.00	426973.00

Dealing with missing values:

In [17]: `df.isna().sum()`

Out[17]:

product_id	0
product_name	0
category	0
discounted_price	0
actual_price	0
discount_percentage	0
rating	0
rating_count	0
about_product	0
user_id	0
user_name	0
review_id	0
review_title	0
review_content	0
img_link	0
product_link	0
dtype: int64	

Observation:

No missing values found

Handling Duplicates:

```
In [20]: # Checking for duplicate rows  
print(df.duplicated().any())
```

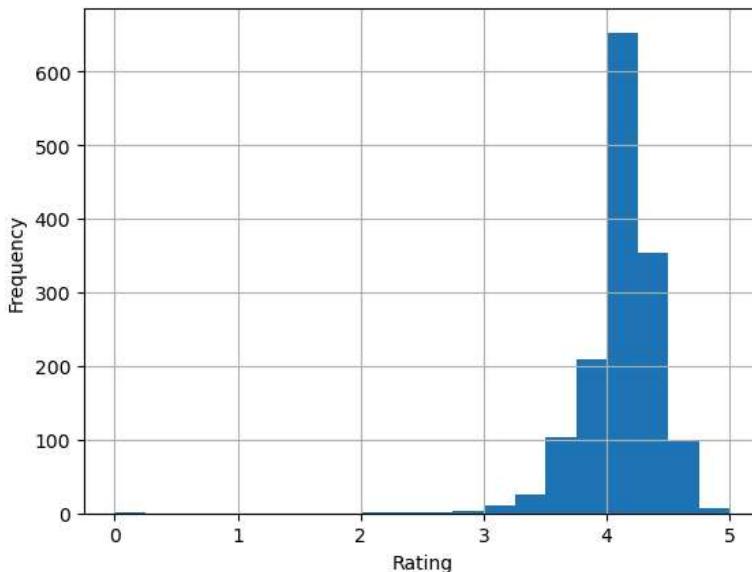
False

Observation:

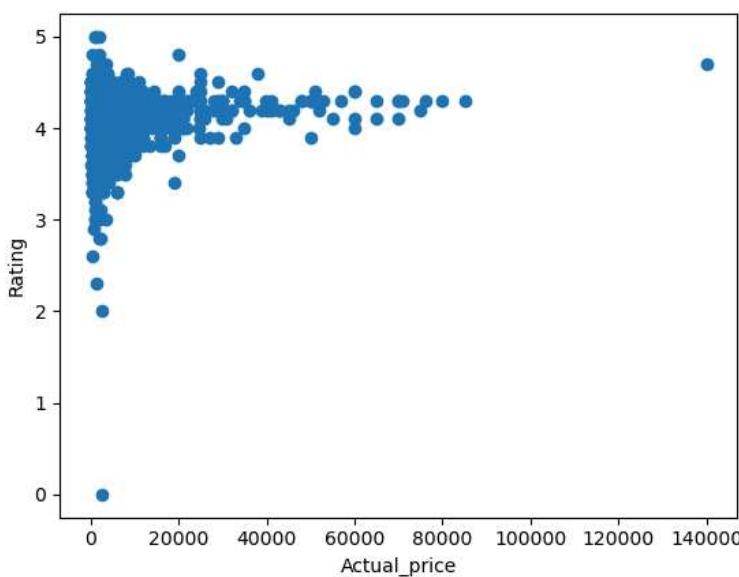
No duplicate values found

Visualisations:

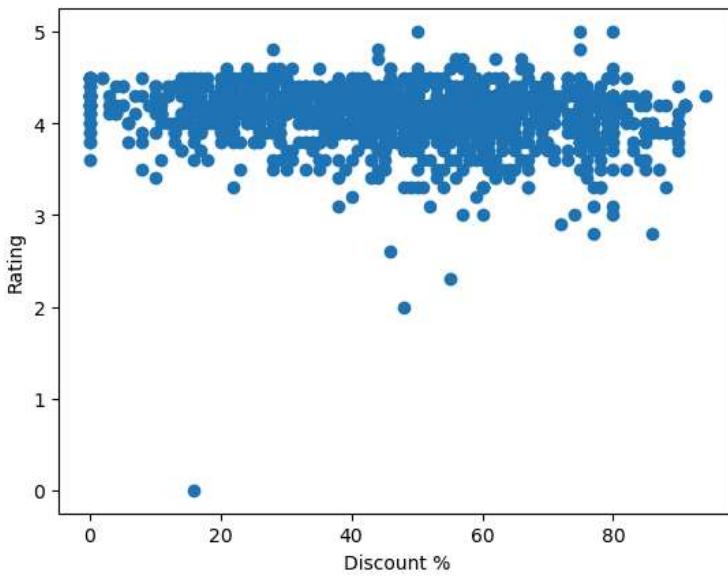
```
In [23]: # Ratings vs their count  
df['rating'].hist(bins=20)  
plt.xlabel('Rating')  
plt.ylabel('Frequency')  
plt.show()
```



```
In [24]: # Plot actual_price vs. rating  
plt.scatter(df['actual_price'], df['rating'])  
plt.xlabel('Actual_price')  
plt.ylabel('Rating')  
plt.show()
```

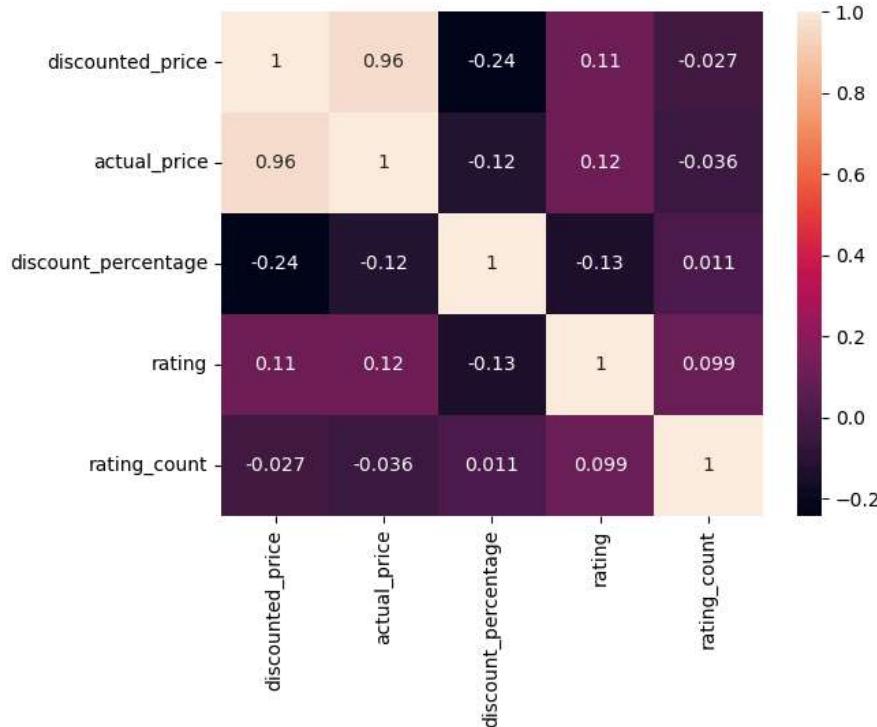


```
In [25]: # Plot discount_percentage vs. rating  
plt.scatter(df['discount_percentage'], df['rating'])  
plt.xlabel('Discount %')  
plt.ylabel('Rating')  
plt.show()
```



```
In [26]: # Plot correlations between numeric variables
sns.heatmap(df[numeric_columns].corr(), annot=True)
```

```
Out[26]: <Axes: >
```



EDA (Exploratory Data Analysis):

```
In [28]: # Top 10 most popular products
pd.set_option('display.max_colwidth', None) # To display full length of content of columns

top_products = df[['product_name', 'rating_count', 'product_link']]\
    .sort_values(by='rating_count', ascending=False).head(10)
print(top_products)
```

```

product_name \
12      AmazonBasics Flexible Premium HDMI Cable (Black, 4K@60Hz, 18Gbps), 3-Foot
65      Amazon Basics High-Speed HDMI Cable, 6 Feet (2-Pack),Black
47      Amazon Basics High-Speed HDMI Cable, 6 Feet - Supports Ethernet, 3D, 4K video,Black
684     AmazonBasics Flexible Premium HDMI Cable (Black, 4K@60Hz, 18Gbps), 3-Foot
400     boAt Bassheads 100 in Ear Wired Earphones with Mic(Furious Red)
352     boAt Bassheads 100 in Ear Wired Earphones with Mic(Taffy Pink)
584     boAt BassHeads 100 in-Ear Wired Headphones with Mic (Black)
370     Redmi 9 Activ (Carbon Black, 4GB RAM, 64GB Storage) | Octa-core Helio G35 | 5000 mAh Battery
371     Redmi 9A Sport (Coral Green, 2GB RAM, 32GB Storage) | 2GHz Octa-core Helio G25 Processor | 5000 mAh Battery
473     Redmi 9A Sport (Carbon Black, 2GB RAM, 32GB Storage) | 2GHz Octa-core Helio G25 Processor | 5000 mAh Battery

rating_count \
12      426973.0
65      426973.0
47      426973.0
684     426972.0
400     363713.0
352     363713.0
584     363711.0
370     313836.0
371     313836.0
473     313832.0

product_link
k
12      https://www.amazon.in/AmazonBasics-Flexible-HDMI-Cable-3-Foot/dp/B07KSMBL2H/ref=sr_1_14?qid=1672909124&s=electronics&sr=1-1
4
65      https://www.amazon.in/AmazonBasics-High-Speed-Cable-2-Pack-Black/dp/B014I8SX4Y/ref=sr_1_73?qid=1672909128&s=electronics&sr=1-7
3
47      https://www.amazon.in/AmazonBasics-High-Speed-HDMI-Cable-Feet/dp/B014I8SSD0/ref=sr_1_51?qid=1672909126&s=electronics&sr=1-5
1
684     https://www.amazon.in/AmazonBasics-Flexible-HDMI-Cable-3-Foot/dp/B07KSMBL2H/ref=sr_1_107?qid=1672903000&s=computers&sr=1-10
7
400     https://www.amazon.in/Boat-BassHeads-100-Inspired-Earphones/dp/B07GQD4K6L/ref=sr_1_69?qid=1672895762&s=electronics&sr=1-6
9
352     https://www.amazon.in/Boat-BassHeads-100-Inspired-Earphones/dp/B07GPXXNNG/ref=sr_1_21?qid=1672895748&s=electronics&sr=1-2
1
584     https://www.amazon.in/boAt-BassHeads-100-Headphones-Black/dp/B071Z8M4KX/ref=sr_1_1?qid=1672902995&s=computers&sr=1-1
1
370     https://www.amazon.in/Redmi-Activ-Carbon-Black-Storage/dp/B09GFVVD9Y/ref=sr_1_39?qid=1672895755&s=electronics&sr=1-3
9
371     https://www.amazon.in/Redmi-9A-Sport-Octa-core-Processor/dp/B09GFLXVH9/ref=sr_1_40?qid=1672895755&s=electronics&sr=1-4
0
473     https://www.amazon.in/Redmi-9A-Sport-Octa-core-Processor/dp/B09GFM8CGS/ref=sr_1_144?qid=1672895784&s=electronics&sr=1-14
4

```

In [29]: # Most popular categories
`df['category'].value_counts()`

Out[29]: category
Computers&Accessories|Accessories&Peripherals|Cables&Accessories|Cables|USBCables 233
Electronics|WearableTechnology|SmartWatches 76
Electronics|Mobiles&Accessories|Smartphones&BasicMobiles|Smartphones 68
Electronics|HomeTheater,TV&Video|Televisions|SmartTelevisions 63
Electronics|Headphones,Earbuds&Accessories|Headphones|In-Ear 52
...
Electronics|Cameras&Photography|Accessories|Batteries&Chargers|BatteryChargers 1
Computers&Accessories|NetworkingDevices|DataCards&Dongles 1
Electronics|HomeAudio|Speakers|MultimediaSpeakerSystems 1
OfficeProducts|OfficePaperProducts|Paper|Copy&PrintingPaper|ColouredPaper 1
Home&Kitchen|Kitchen&HomeAppliances|Vacuum,Cleaning&Ironing|Vacuums&FloorCare|VacuumAccessories|VacuumBags|HandheldBags 1
Name: count, Length: 211, dtype: int64

In [30]: # Extract the main category
`df['main_category'] = df['category'].str.split('|').str[0]`
`print(df[['category', 'main_category']].sample(5))`
`df['main_category'].value_counts()`

```

category \
1355 Home&Kitchen|Kitchen&HomeAppliances|WaterPurifiers&Accessories|WaterPurifierAccessories
1360                               Home&Kitchen|Heating,Cooling&AirQuality|RoomHeaters|FanHeaters
128                                Electronics|HomeTheater,TV&Video|Accessories|RemoteControls
1210 Home&Kitchen|Kitchen&HomeAppliances|WaterPurifiers&Accessories|WaterFilters&Purifiers
342      Electronics|Mobiles&Accessories|Smartphones&BasicMobiles|Smartphones

main_category
1355 Home&Kitchen
1360 Home&Kitchen
128   Electronics
1210 Home&Kitchen
342   Electronics

Out[30]: main_category
Electronics      526
Computers&Accessories    453
Home&Kitchen     448
OfficeProducts    31
MusicalInstruments  2
HomeImprovement    2
Toys&Games        1
Car&Motorbike      1
Health&PersonalCare  1
Name: count, dtype: int64

```

```

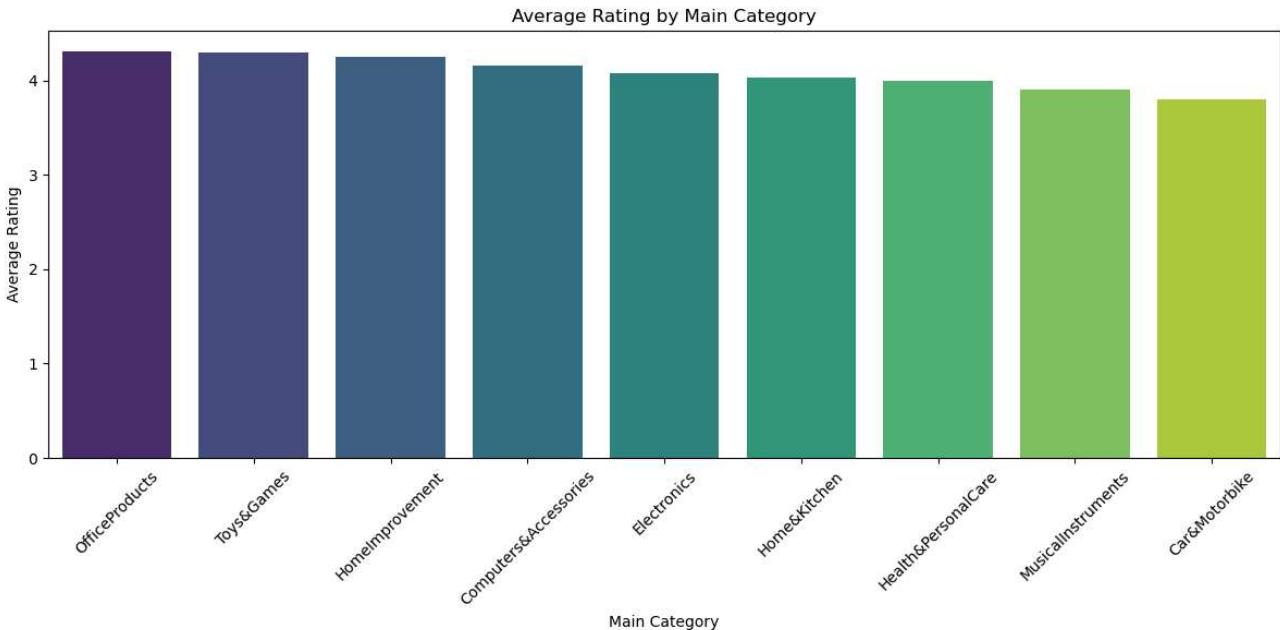
In [31]: # Calculate average rating per main_category
avg_ratings = df.groupby('main_category')['rating'].mean().reset_index()

# Sort by rating
avg_ratings = avg_ratings.sort_values(by='rating', ascending=False)

# Create bar plot
plt.figure(figsize=(12, 6))
sns.barplot(data=avg_ratings, x='main_category', y='rating', palette='viridis')

# Customize plot
plt.xticks(rotation=45)
plt.title('Average Rating by Main Category')
plt.xlabel('Main Category')
plt.ylabel('Average Rating')
plt.tight_layout()
plt.show()

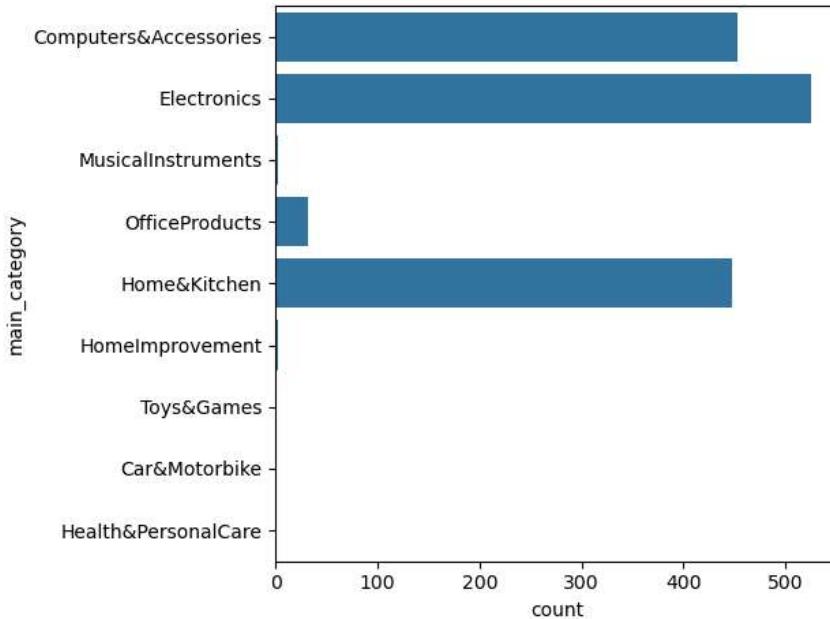
```



```

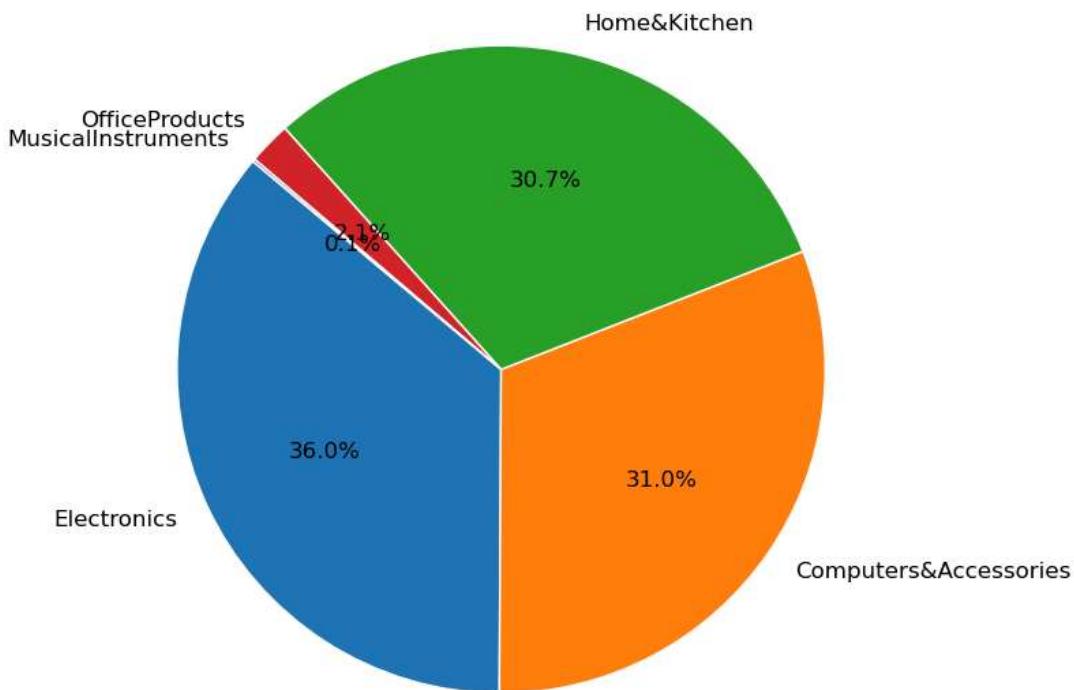
In [32]: # Number of products per category
sns.countplot(y='main_category', data=df)
plt.tight_layout()
plt.show()

```



```
In [33]: # Plot the proportion of top 5 categories as a pie chart
plt.figure(figsize=(8, 8)) # Increase the figure size
df['main_category'].value_counts().head(5).plot.pie(
    autopct='%1.1f%%',
    startangle=140,
    wedgeprops={'edgecolor': 'white'},
    textprops={'fontsize': 12}
)
plt.ylabel('') # Remove y-axis label
plt.title('Top 5 Main Categories by Proportion')
plt.tight_layout()
plt.show()
```

Top 5 Main Categories by Proportion



Sentiment Analysis:

Method 1: Using NLTK VADER (Valence Aware Dictionary for sEntiment Reasoning)

Sentiment Analysis using VADER (Valence Aware Dictionary for sEntiment Reasoning) is a rule-based method for determining the emotional tone (positive, negative, neutral) of text, particularly designed to work well on social media, short texts, and informal language.

Key points about VADER:

Lexicon-based: Uses a dictionary where each word is assigned a sentiment score between -4 (most negative) and +4 (most positive).

Handles social media slang: Understands emoticons (:-)), acronyms (LOL), and slang words.

Accounts for context: Adjusts sentiment based on punctuation, capitalization, degree modifiers (e.g., very, extremely).

Outputs four scores:

pos: proportion of positive sentiment

neu: proportion of neutral sentiment

neg: proportion of negative sentiment

compound: a normalized score between -1 (most negative) and +1 (most positive)

```
In [36]: df['review_content']
```

Out[36]: 0

Looks durable Charging is fine tooNo complains,Charging is really fast, good product.,Till now satisfied with the quality.,This is a good product . The charging speed is slower than the original iPhone cable,Good quality, would recommend,https://m.media-amazon.com/images/W/WEBP_402378-T1/images/I/81---F1ZgHL._SY88.jpg,Product had worked well till date and was having no issue.Cable is also sturdy enough...Have asked for replacement and company is doing the same...,Value for money

1

I ordered this cable to connect my phone to Android Auto of car. The cable is really strong and the connection ports are really well made. I already has a Micro USB cable from Ambrane and it's still in good shape. I connected my phone to the car using the cable and it got connected well and no issues. I also connected it to the charging port and yes it has Fast Charging support.,It quality is good at this price and the main thing is that i didn't ever thought that this cable will be so long it's good one and charging power is too good and also supports fast charging,Value for money, with extra length👉,Good, working fine,Product quality is good,Good,very good,Bought for my daughter's old phone.Brand new cable it was not charging, I already repacked and requested for replacement.I checked again, and there was some green colour paste/fungus inside the micro USB connector. I cleaned with an alcoholic c and starts working again.Checked the ampere of charging speed got around 1400ma-1500ma - not bad, came with braided 1.5m long cable, pretty impressive for the price.Can't blame the manufacturer.But quality issues by the distributor, they might have stored in very humid place.

2

Not quite durable and sturdy,https://m.media-amazon.com/images/W/WEBP_402378-T1/images/I/71rIggrbUCL._SY88.jpg,Working good,https://m.media-amazon.com/images/W/WEBP_402378-T1/images/I/61bkp9Y06wl._SY88.jpg,Product,Very nice product,Working well,It's a really nice product

3

Good product,long wire,Charges good,Nice,I bought this cable for Rs.339 worthy product for this price, i tested it in various charger adapters 33w and 18w it supports fast charging as well.,Good,Ok,I had got this at good price on sale on Amazon and product is useful with warranty but for warranty you need to go very far not practical for such a cost and mine micro to type c connector stopped working after few days only.,I like this product

4 Bought this instead of original apple, does the work for 150rs, not as fast as apple charger but its a good option if you want cheap and good product, bought it for iPad pro 10.5 and it's working flawlessly, build quality is ok, its not like i am gonna hang my clothes on it and i want a very strong cable, even a braided cable stop working after a year, i have used both Anker and Apple store strong braided cable they all stop working after a year so please don't buy high end cables just for that instead choose a this one and even if it's stops working withing a year you only loose 150rs compares to 2000rs.UPDATE-----pin has stopped charging from one side, now i have to slip the pin to charge from other side, but i will update and let know for how long does it work,,It's good. Not sure about durability as the pin area feels a bit fragile,Does not support apple ca rplaySo was little disappointed about thatOther than that cable is made up of very good quality,Best to buy,100% NOT FATHFUL,Writing this review post 10 months and 3 orders of the same product.Honestly Portronics Konnect L lightning cable works like magic with the original Apple charging brick.Seeing the price of the cable I initially hesitated buying as it was as low as ₹99/- with the offers and so I wasn't sure if it would work well with my iPhone 12 or whether it would impact my iPhone's battery health because all the other lightning cable brands were costing over ₹350/- like Wayona, Amazon Basics, etc.Earlier I was using Wayona Brand lightning cable with eventually frayed and stopped working.Charging Speed:Charges my iPhone fast enough almost similar compared to the original cable level when used with 12W Original Apple power adapter.Quality and Durability:Great quality braided cable and does n't tangle easily and can withstand day-to-day usage.L-Shaped pin:This is very innovative by Portronics and it makes sure the cable doesn't get damaged even if used while charging.CarPlay and Data Sync:Works smoothly with CarPlay and syncs data effortlessly.P S: I have used this cable only with the Original Apple Charging Brick and extremely satisfied with its performance.,Better than I expect the product I like that Quality and I plan to buy same type cable come with usb C to Lighting cable for emergency purpose that much I love this cable. Buy for this cable only emergency uses only since Good one,Good product and value for money

...

1460

I received product without spanner,Excellent product,It's little spongy compared to my last purchased brand which was hard and not so spongy as this one.,Very good and easy to install,Comes in separate sealed packets & a spanner for easy installation.fitting was perfect.. really nice product,seem to be ok product.,The candels are okay not quite good. For this thickness we need to change for every 2 months or sometimes even every one month,Price should be more little bit less

1461

ok,got everything as mentioned but the measuring cup was broken didn't expect this from amazon and as well as prestige. except this everything was good,I had a bad experience buying this Prestige 1L rice cooker. The screw on the bottom was not placed properly. There was a gap between the bottom plate and the top portion. I regret buying this. they already had delivered the replacement but it came worst than the other one so I returned the product.,Like it for less power consumption & quick preparations.,The product is ok but it is small more small than i expected. It's perfect for 1 or2 people. Even my water bottle is taller.,The product is used for cooking in hostel. Only use sponges to clean the utensil that comes with the cooker. Don't modify it.Use of excess water spoils the work. You can saute. Vegetables for two minutes also. Ideal for students living in hostel.,i'm using from two months. it is nice and no worries about product quality and strongly recommended for to cook for two or three persons,Very Good product

1462

plastic but cool body ,u have to find sturdy surface to put it vertically , on plastic stool it vibrate in vertical position ,u need bigger plug point for it ,very fast and effective heating ,it through cold air too ,price is bit high but it's bajaj so good.very good buy.,Using after only 2 months, the front part of this bajaj rx10, room heater melt starting.,Useful in winter, box gets heated which may cause damage but so far so good.,The body flaps look very delicate and did not align well when received. I had to fix them,This Heater is good for a room with 10X10 ft dimension. The best part is 2 years warranty from Bajaj. Personally recommended you to purchase this product to stay stress free.,RECEIVED DEFECTIVE PRODUCT FROM AMAZON SELLER M/S ROYAL ELECTRONICS INDIA.I AM SURPRISED ,HOW AMAZON IS ALLOWING THIS TYPE OF SELLERS ON THEIR PLATFORM.,Nice product,Nice product

1463

I have installed this in my kitchen working fine is just fan speed is very slow could have been faster but not it is slow thats the reason this exhaust smoke very slow but it does exhaust.,Good quality,https://m.media-amazon.com/images/W/WEBP_402378-T1/images/I/61izko3RxDL._SY88.jpg,speed could have been little bit more,,Beat performance this product.,Connecting wire is too short,Super 1464

It does it job perfectly..only issue is temp control is not perfect . U need to keep checking the brownness of sandwich untill it gets upto ur choise,Product is good, but the control knob is to be modified cause after the cooking process completes the control knob should be return back to initial starting place. If not that process timer should be given basically. So think well what happens whether didn't had atleast one option.Not a indicator option.,It is a nice product, easy to use. The grilling function is also wonderful. But the exterior metallic finish picks up finger marks and other spots which tend to become permanent. Cleaning them is almost impossible.Very good product,This is a pretty powerful sandwich maker, for home use. The product looks good, has a sturdy build, and heats up quick. The sandwiches are made pretty quickly, without burning.,बोरोसिल ब्रॉड का यह "सेंडविच मेकर" देखने में तो अच्छा लगता है मगर इसकी बिल्ड कालिटी अच्छी नहीं है। यह लगभग Rs 3000 के आसपास आता है। इस प्रकार की बिल्ड कालिटी ₹2000 के करीब मिल जाती है तो कोई क्यों ₹1000 अंधिक भुगतान करें। पहले मैंने इसे review देखने के बाद ऑर्डर किया था लेकिन जब घर पर डिलीवरी होने के पश्चात unboxing करके देखा तो इस

की बिल्ड कालिटी कुछ खास नहीं लगी इसलिए अपने पैसे बचाने के लिए मैंने इसे वापस भेज दिया। मैंने इसकी पैकेजिंग, मैनुअल और सेंडविच भेकर की फोटोग्राफ आप सभी से शेयर की है। आप स्थय देख के अनुमान लगा सकते हैं।, Recommend work as expected, Its easy tp use
Name: review_content, Length: 1465, dtype: object

In [37]: df['review_title']

Out[37]: 0
Satisfied,Charging is really fast,Value for money,Product review,Good quality,Good product,Good Product,As of now seems good
1
A Good Braided Cable for Your Type C Device,Good quality product from ambrane,Super cable,As,Good quality,Good product,its good,Good quality for the price but one issue with my unit
2
Good speed for earlier versions,Good Product,Working good,Good for the price,Good,Worth for money,Working nice,it's a really nice product
3
Good product,Good one,Nice,Really nice product,Very first time change,Good,Fine product but could be better,Very nice it's charging like jet
4
As good as original,Decent,Good one for secondary use,Best quality,GOOD,Amazing product at a mind blowing price!,Nice Quality,Good product

...
1460
Received the product without spanner,Excellent product,Satisfactory,Good product,great product,performance yet to be checked?,Value for money,Good product
1461 ok,everything was good couldn't return bcoz I needed it in emergency,Don't buy this 1 litre Prestige rice cooker. Already replaced but still got damaged product,Ideal for a small family of two.,Nice but small,Read the leaflets before you start cooking. Learn to use optimum quantity of water.,Quality excellent,Very Good product
1462
very good,Work but front melt after 2 month,Good one,It is durable,Review.,DEFECTIVE PRODUCT,Nice product,Nice product
1463
Fan Speed is slow,Good quality,Good product,good,Old is gold.,Good product,Nice product,Super ❤️
1464 Works perfectly,Ok good product,Nice Product. Recommend it. But cleaning its exterior is cumbersome.,Excellent product 🌟,A good product for household use,मुझे बिल्कुल भी मजा नहीं आया और वापस कर दिया।,Best product,Good
Name: review_title, Length: 1465, dtype: object

In [38]: # Download necessary NLTK resources

```
# nltk.download('stopwords')
# nltk.download('punkt')
# nltk.download('vader_lexicon')
# nltk.download('punkt_tab') # Download the punkt_tab resource to avoid the LookupError

# Initialise english stopwords
stop_words = set(stopwords.words('english'))

# Initialize SentimentIntensityAnalyzer
sia = SentimentIntensityAnalyzer()

# Function to get stopwords from the review content
def get_stopwords_from_text(text):
    if isinstance(text, str): # Ensure the text is a string
        words = word_tokenize(text.lower()) # Tokenize and convert to lower case
        stop_words_in_text = [word for word in words if word in stop_words]
        return stop_words_in_text
    return [] # Return empty list if text is not a valid string

# Function to analyze sentiment
def analyze_sentiment(text):
    if isinstance(text, str): # Ensure the text is a string
        score = sia.polarity_scores(str(text)) # Always convert to string to avoid errors
        compound = score['compound']
        if compound >= 0.05:
            return 'positive'
        elif compound <= -0.05:
            return 'negative'
        else:
            return 'neutral'
    return 'neutral' # Return 'Neutral' if text is not a valid string

# Apply the stopwords function and sentiment analysis
df['stopwords'] = df['review_title'].apply(get_stopwords_from_text)
df['sentiment_1'] = df['review_title'].apply(analyze_sentiment)

print(df[['review_title', 'stopwords', 'sentiment_1']])
```

```

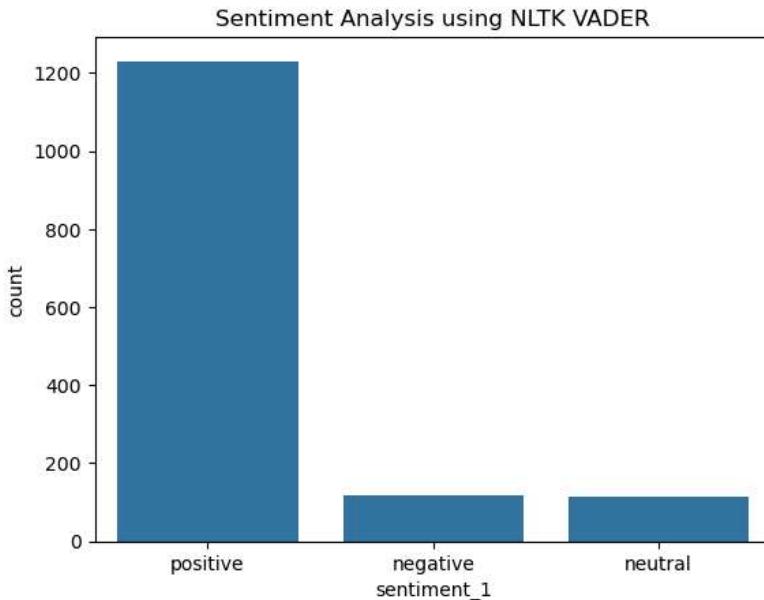
review_title \
0
Satisfied,Charging is really fast,Value for money,Product review,Good quality,Good product,Good Product,As of now seems good
1
A Good Braided Cable for Your Type C Device,Good quality product from ambrane,Super cable,As,Good quality,Good product,its good,Good
d quality for the price but one issue with my unit
2
Good speed for earlier versions,Good Product,Working good,Good for the price,Good,Worth for money,Working nice,it's a really nice p
roduct
3
Good product,Good one,Nice,Really nice product,Very first time change,Good,Fine product but could be better,Very nice it's charging
like jet
4
As good as original,Decent,Good one for secondary use,Best quality,GOOD,Amazing product at a mind blowing price!,Nice Quality,Good
product
...
...
1460
Received the product without spanner,Excellent product,Satisfactory,Good product,great product,performance yet to be checked?,Value
for money,Good product
1461 ok,everything was good couldn't return bcoz I needed it in emergency,Don't buy this 1 litre Prestige rice cooker. Already rep
laced but still got damaged product,Ideal for a small family of two.,Nice but small,Read the leaflets before you start cooking. Lea
rn to use optimum quantity of water.,Quality excellent,Very Good product
1462
very good,Work but front melt after 2 month,Good one,It is durable,Review.,DEFECTIVE PRODUCT,Nice product,Nice product
1463
Fan Speed is slow,Good quality,Good product,good,Old is gold.,Good product,Nice product,Super ❤️
1464
Fan Speed is slow,Good quality,Good product,good,Old is gold.,Good product,Nice product,Super ❤️ Works perfe
ct,Ok good product,Nice Product. Recommend it. But cleaning its exterior is cumbersome.,Excellent product 🌟,A good product for hous
ehold use,मुझे बिल्कुल भी मजा नहीं आया और वापस कर दिया।,Best product,Good
stopwords \
0
[is, for, as, of, now]
1
[a, for, your, from, as, its, for, the, but, with, my]
2
[for, for, the, for, it, a]
3
[very, but, be, very, it]
4
[as, as, for, at, a]
...
1460
[the, to, be, for]
1461 [was, i, it, in, do, this, but, for, a, of, but, the, before, you, to, of, very]
1462
[very, but, after, it, is]
1463
[is, is]
1464
[it, but, its, is, a, for]

sentiment_1
0
positive
1
positive
2
positive
3
positive
4
positive
...
1460
neutral
1461
positive
1462
neutral
1463
positive
1464
positive

```

[1465 rows x 3 columns]

```
In [39]: # Plot Sentiment Distribution
sns.countplot(x='sentiment_1', data=df)
plt.title("Sentiment Analysis using NLTK VADER")
plt.show()
```



```
In [40]: # Combine texts by sentiment
all_reviews = ' '.join(df['review_content'].astype(str)) # all_reviews is a single string containing all the review texts in the df
positive_reviews = ' '.join(df[df['sentiment_1'] == 'positive']['review_title'].astype(str))
neutral_reviews = ' '.join(df[df['sentiment_1'] == 'neutral']['review_title'].astype(str))
negative_reviews = ' '.join(df[df['sentiment_1'] == 'negative']['review_title'].astype(str))

# Generate word clouds
wordclouds = {
    'All Reviews': WordCloud(width=800, height=400, background_color='white', colormap='viridis').generate(all_reviews),
    'Positive Reviews': WordCloud(width=800, height=400, background_color='white', colormap='Greens').generate(positive_reviews),
    'Neutral Reviews': WordCloud(width=800, height=400, background_color='white', colormap='gray').generate(neutral_reviews),
    'Negative Reviews': WordCloud(width=800, height=400, background_color='white', colormap='Reds').generate(negative_reviews),
}

# Plot in subplots
fig, axs = plt.subplots(4, 1, figsize=(14, 20))

for i, (title, wc) in enumerate(wordclouds.items()):
    axs[i].imshow(wc, interpolation='bilinear')
    axs[i].set_title(title, fontsize=18)
    axs[i].axis('off')

plt.tight_layout(pad=5.0)
plt.show()
```



Positive Reviews



Neutral Reviews





Method 2 : Using deep learning model: HuggingFace Transformers (BERT-based Roberta pretrained model)

Sentiment Analysis using Deep Learning Models (Hugging Face Transformers, e.g., BERT-based RoBERTa) is a data-driven approach where a pre-trained transformer model understands the meaning of text and predicts its sentiment (positive, negative, neutral, etc.) based on learned patterns from large datasets.

Key Concepts-

Transformer Architecture: Uses self-attention mechanisms to understand word meaning in context (e.g., "bank" in "river bank" vs "money bank").

RoBERTa Model: Stands for Robustly Optimized BERT Approach. Trained on more data and with better hyperparameters than BERT for higher accuracy. Available as pre-trained weights in Hugging Face’s transformers library.

Transfer Learning: The pre-trained model is fine-tuned on sentiment-labeled data (e.g., movie reviews, tweets) to classify text into sentiment categories.

Advantages over VADER:

Captures nuance and context better.

Works well with longer, complex text.

Can adapt to domain-specific language with fine-tuning.

```
In [42]: # Load model and tokenizer
model_name = "cardiffnlp/twitter-roberta-base-sentiment"
tokenizer = AutoTokenizer.from_pretrained(model_name)
model = AutoModelForSequenceClassification.from_pretrained(model_name)

# Define Label mapping (based on model order)
labels = ['negative', 'neutral', 'positive']

# Function to get the sentiment label with highest probability
def get_sentiment_label(text):
    encoded_input = tokenizer(str(text), return_tensors='pt', truncation=True)
    with torch.no_grad():
        output = model(**encoded_input)
    scores = output.logits[0]
    probabilities = torch.nn.functional.softmax(scores, dim=0)
    predicted_class = torch.argmax(probabilities).item()
    return labels[predicted_class]

# Apply on the DataFrame
df['sentiment_2'] = df['review_title'].apply(get_sentiment_label)

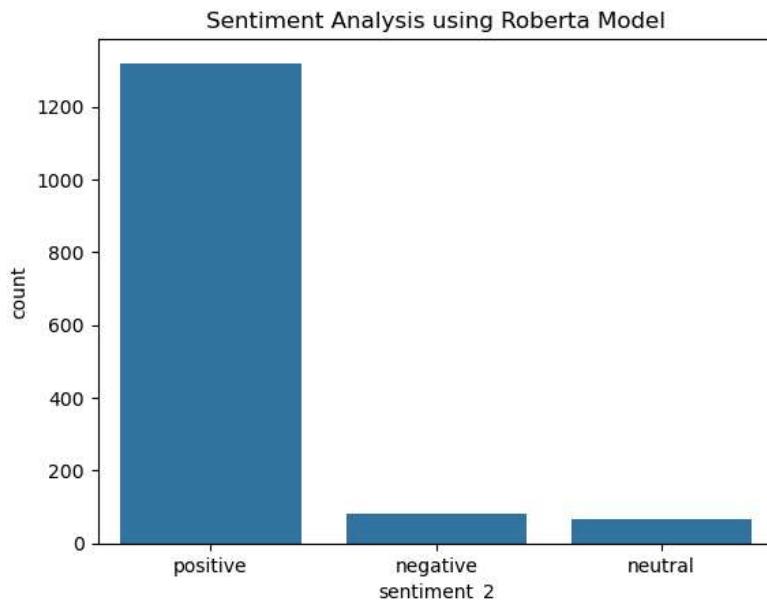
# Preview
print(df[['review_title', 'sentiment_2']].head())
```

```
Asking to truncate to max_length but no maximum length is provided and the model has no predefined maximum length. Default to no truncation.
```

```
review_title \
0                                     Satisfied,Charging is really fast,Value for money,Product review,Good quality,Good product,Good Product,As of now seems good
1 A Good Braided Cable for Your Type C Device,Good quality product from ambrane,Super cable,As,Good quality,Good product,its good,Good quality for the price but one issue with my unit
2                                     Good speed for earlier versions,Good Product,Working good,Good for the price,Good, Worth for money,Working nice,it's a really nice product
3                                     Good product,Good one,Nice,Really nice product,Very first time change,Good,Fine product but could be better,Very nice it's charging like jet
4                                     As good as original,Decent,Good one for secondary use,Best quality,GOOD,Amazing product at a mind blowing price!,Nice Quality,Good product

sentiment_2
0   positive
1   positive
2   positive
3   positive
4   positive
```

```
In [43]: # Plot Sentiment Distribution
sns.countplot(x='sentiment_2', data=df)
plt.title("Sentiment Analysis using Roberta Model")
plt.show()
```



```
In [44]: print("Sentiment 1 counts (VADER):")
print(df['sentiment_1'].value_counts())
print()

print("Sentiment 2 counts (RoBERTa):")
print(df['sentiment_2'].value_counts())
```

```
Sentiment 1 counts (VADER):
sentiment_1
positive    1231
negative     119
neutral      115
Name: count, dtype: int64

Sentiment 2 counts (RoBERTa):
sentiment_2
positive    1320
negative     80
neutral      65
Name: count, dtype: int64
```

```
In [45]: # Count sentiment labels
sentiment1_counts = df['sentiment_1'].value_counts().reindex(['negative', 'neutral', 'positive'], fill_value=0)
sentiment2_counts = df['sentiment_2'].value_counts().reindex(['negative', 'neutral', 'positive'], fill_value=0)

# Labels and positions
labels = ['Negative', 'Neutral', 'Positive']
x = np.arange(len(labels)) # Label Locations
```

```

width = 0.35 # width of the bars

# Plot
fig, ax = plt.subplots(figsize=(8, 6))
bars1 = ax.bar(x - width/2, sentiment1_counts, width, label='Sentiment 1 (VADER)', color="#1f77b4")
bars2 = ax.bar(x + width/2, sentiment2_counts, width, label='Sentiment 2 (RoBERTa)', color="#ff7f0e")

# Labels & title
ax.set_ylabel('Count')
ax.set_title('Sentiment Label Counts: VADER vs RoBERTa')
ax.set_xticks(x)
ax.set_xticklabels(labels)
ax.legend()

# Annotate bars
for bar in bars1 + bars2:
    height = bar.get_height()
    ax.annotate(f'{int(height)}',
                xy=(bar.get_x() + bar.get_width() / 2, height),
                xytext=(0, 3), # vertical offset
                textcoords="offset points",
                ha='center', va='bottom', fontsize=9)

plt.tight_layout()
plt.show()

```

