

Table of Contents

1 Introduction	2
2 Methodologies.....	2
2.1 Data Preparation	2
2.2 Aspect-Opinion Extraction (AOE)	3
2.3 Semantic Search	3
2.4 Sentiment Analysis (SA).....	4
3 Evaluation	4
3.1 Manual check	4
3.2 Prediction model.....	4
4 Analyses and Recommendation	5
4.1 Feature improvement	5
4.2 Competitive landscape	5
5 Limitations.....	7
6 Conclusion	7
Appendices	9
Appendix 1 – Manual Evaluation Result.....	9
Appendix 2 – Chat GPT prompts.....	10
References.....	71

1 Introduction

In e-commerce, customer feedback shapes product development and marketing strategies (Kolasani, 2023). The importance of understanding sentiments in reviews motivated us to employ Aspect-Based Sentiment Analysis (ABSA), a technique providing granularity in assessing specific attributes (D'Aniello et al., 2022). This study focuses on sentiment analysis (SA) of the latest smartphones from top brands on Amazon, extracting sentiments for key aspects: camera, battery, display, operation, price, and design. Derived insights aid in feature enhancement, addressing pain points, and customizing marketing strategies to consumer preferences (Kumar et al., 2024).

2 Methodologies

Our approach leveraged various well-trained models and packages with the procedures illustrated in Figure 1.

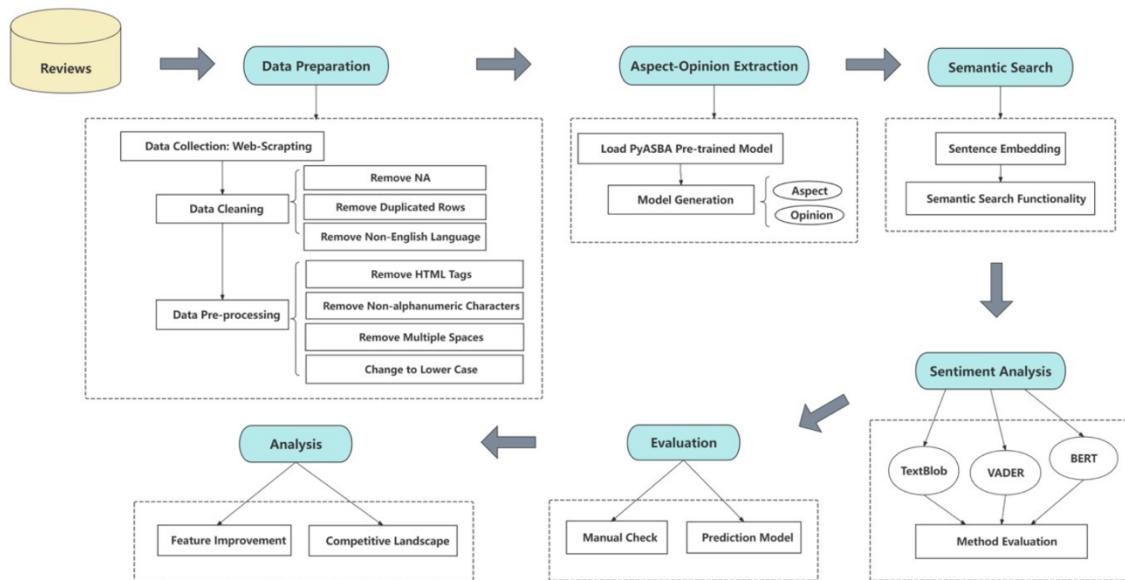


Figure 1: Project Workflow

2.1 Data Preparation

Data preparation comprises data collection and processing. Here, we leveraged Selenium and BeautifulSoup to automate web scraping on Amazon to gather real-time smartphone information (name, brand, price) and review details (title, rating, date, text) from top four brands (Thomas and Martur, 2019).

Subsequently, we conducted data preprocessing by removing 'N/A'; filtering English reviews; eliminating duplicates; removing HTML tags, non-alphanumeric characters, and extra spaces.

Although our ABSA model can deal with multilingual tasks, English reviews were selected to simplify the subsequent steps. Our final dataset consisted of 593 unique observations.

Table 1: Data Description

Website	Brand	Time	Number of Reviews
Amazon.com	Apple	Jun 2021 – Apr 2024	142
Amazon.com	One Plus	Dec 2018 – Apr 2024	150
Amazon.com	Samsung	Jul 2023 – Apr 2024	150
Amazon.com	Huawei	Oct 2018 – Apr 2024	151

2.2 Aspect-Opinion Extraction (AOE)

We used the ABSA-Instruction model from PyABSA package, a modularised framework built on PyTorch to extract the aspects, corresponding opinions, and polarity from each review (Yang et al., 2023). This pre-trained model is fine-tuned on the InstructABSA, an instruction-tuned T5-based model for aspect term extraction (Scaria et al., 2023). Since this model is trained on the Laptop14 dataset (Pontiki et al., 2014) comprising laptop reviews, it has relatively similar aspects to our domain, smartphones.

Based on the results, four important attributes including aspect, polarity, opinion, and category were extracted into new columns. We also cleaned the new data by splitting multi-valued rows into single-valued ones and adjusting the format for category words.

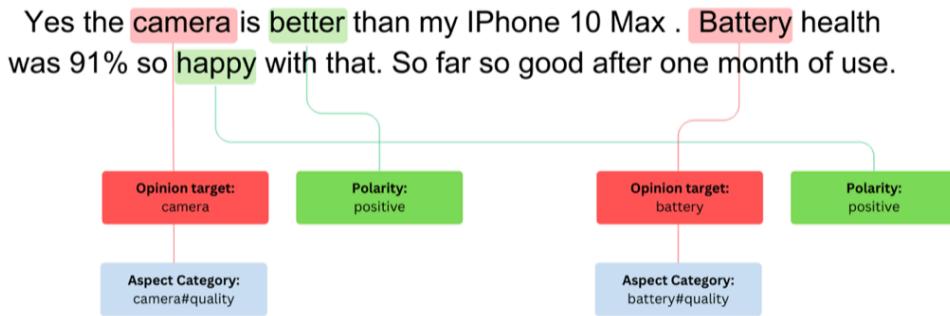


Figure 2: Illustration of Aspect-Opinion Extraction

2.3 Semantic Search

Next, we utilised semantic search to associate each category with the closest predefined aspects (i.e. camera, battery, display, operation, price, and design), essential smartphone features to customers (Rashid et al., 2020). Instead of using 'aspect' from the ABSA model, we employed 'category', which aligns better with our target aspects. Specifically, 'category'

and ‘predefined aspects’ were embedded by Sentence Embeddings from Sentence Transformer. Subsequently, a search functionality was designed to calculate the cosine similarity between category embeddings and predefined aspect embeddings (Liu et al., 2018). If the highest similarity exceeded a threshold (0.5), it assigned the closest aspect to the category, saved as a new column in the data; otherwise, it labelled the category as “Others”.

2.4 Sentiment Analysis (SA)

Finally, we conducted SA to the opinions of each aspect of every review through three methods: TextBlob, VADER and BERT. After comparing their sentiment scores, we selected VADER because:

- BERT only captures polarity and confidence score, which is not versatile enough to extract the sentiment scores of each aspect of the products.
- VADER is highly tuned for social media and can interpret slang, emojis, and emoticons effectively. Hence, it excels in handling customer reviews compared to static lexicons or simplistic rule-based approaches like TextBlob (Wan Min et al., 2020).
- On our dataset, VADER’s scores aligned closely with the polarity predicted by BERT Sentiment, a fine-tuned model possessing comprehensive language understanding and contextual awareness (Sayeed et al., 2023), affirming the reliability and precision of VADER’s predictions.

3 Evaluation

To validate our ABSA models, we employed both manual check and prediction model approaches.

3.1 Manual check

To ensure the quality of SA, we extracted a small sample, assessing the review texts and comparing with the aspect and the sentiment score/polarity from the models. In 10 random instances, sentiment scores perfectly matched with our manual assessments while 1 aspect was misidentified (refer to [Appendix 1](#)).

3.2 Prediction model

Using Random Forest classifier, we predicted the customer satisfaction level from the review (Y) from sentiment scores of 6 aspects (X). Y, labelled ‘Type’, is derived from the ‘review_rating’ column, categorising reviews as ‘high’ or ‘low’. We divided the data into training

and testing sets (80:20 respectively) and evaluated the model's accuracy using classification metrics (Table 2).

Table 2: Classification Metrics

Type	Accuracy	Precision	Recall	F1-Score
High	94%	95%	98%	97%
Low	94%	75%	50%	60%

The result shows high accuracy (94%) and a slightly low F1-score and Recall for 'Low' review type. This suggested that these sentiment scores are reliable indicators of customer satisfaction, which can be generalised to further predictive models.

4 Analyses and Recommendation

4.1 Feature improvement

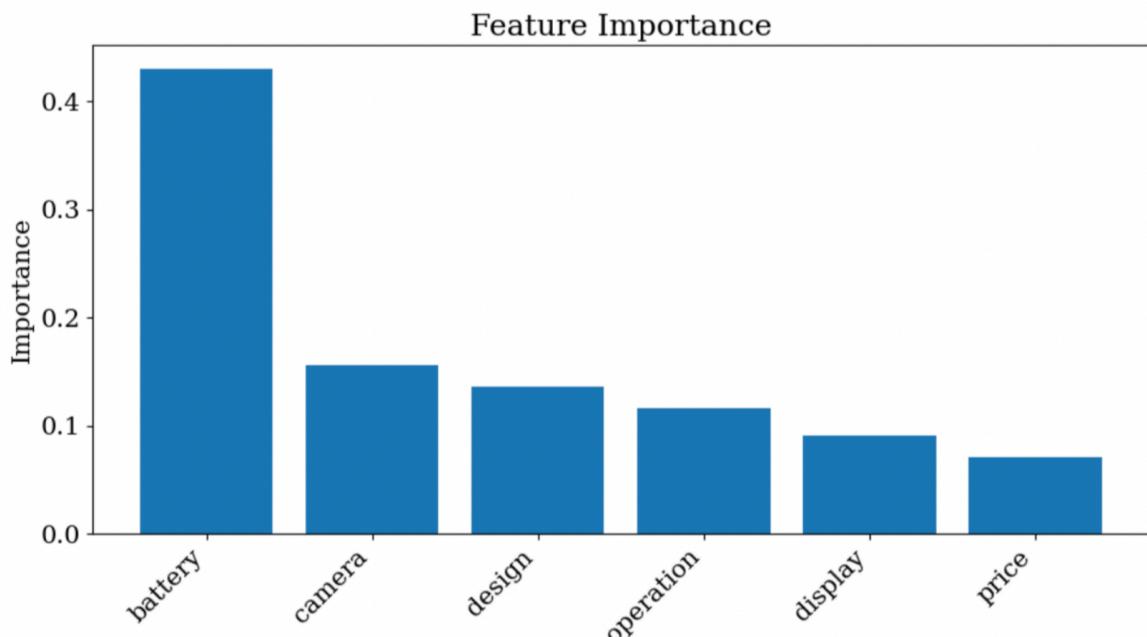


Figure 3: Feature Importance

Battery emerges as the paramount feature for customers, evidenced by its highest importance score from the prediction model (Figure 3). This underscores the significance of prioritising battery performance enhancement in product development. By bolstering battery longevity and efficiency in their devices, companies significantly appeal to their customer base and drive greater satisfaction.

4.2 Competitive landscape

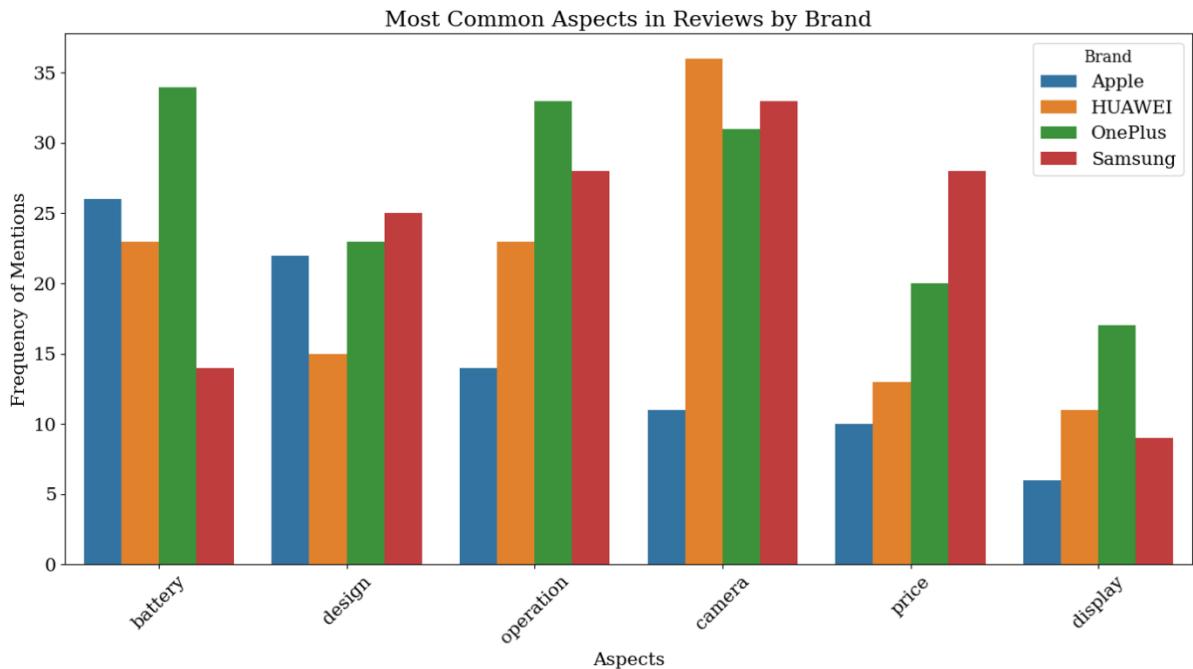


Figure 4: Most Common Aspects in Review by Brand

Figure 4 shows the frequency of mentions for different smartphone aspects across 4 brands, indicating areas of consumer interests. This data can help brands tailor their marketing and product strategies accordingly. For instance, since OnePlus reviews frequently mentioned operations, the company could prioritise promoting processor enhancements to attract more customers.

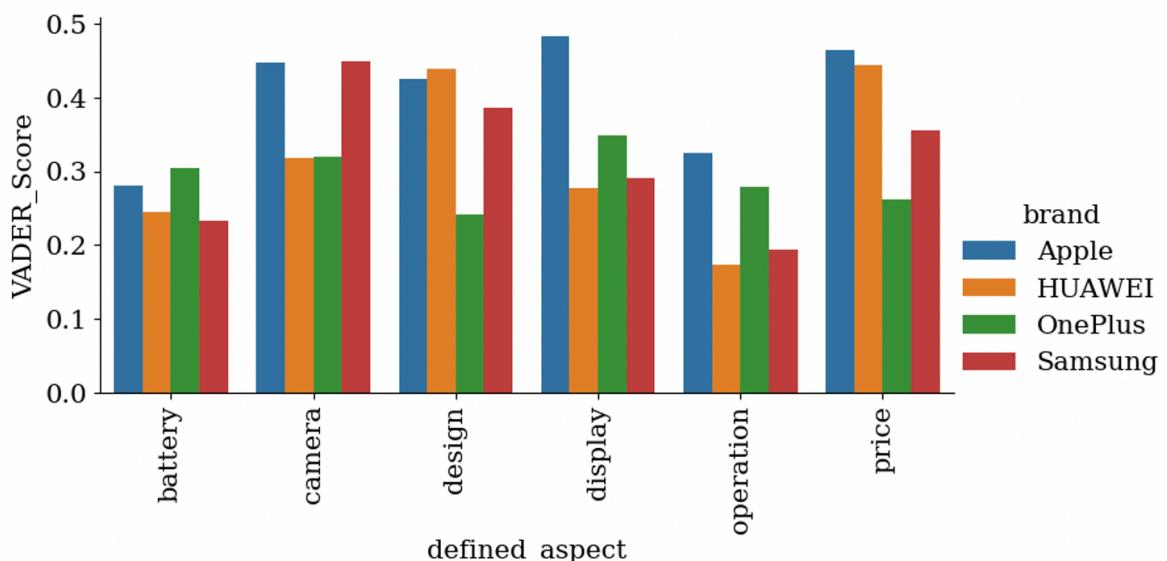


Figure 5: Sentiment Score by Aspects by Brand

Figure 5 highlights brand strengths and weaknesses in 6 key aspects in terms of customer satisfaction. This visualisation enables brands to pinpoint areas needing improvement and to

strategise marketing efforts effectively. High sentiment scores indicate a brand's strengths, suggesting market leadership in those aspects. For instance, Apple's highest score in terms of display underscores its market dominance in screen technology, which could be emphasised in promotional campaigns to attract consumers valuing high-quality visual experiences. Conversely, lower scores, as seen with OnePlus in the display category, indicate areas for improvement, suggesting that enhancements in screen technology could enhance future models and marketing strategies. This analysis provides actionable insights, helping brands align their product development and marketing with consumer expectations.

5 Limitations

We acknowledge several limitations in our work:

1. Rely solely on English Amazon reviews that potentially overlook valuable insights from other online platforms or offline sources and diverse language customers. It is recommended to collect reviews from various sales channels and run our ABSA methodology on multilingual data.
2. Our focus on specific attributes may overlook other crucial factors influencing consumer satisfaction, possibly resulting in incomplete insights.
3. Our models misidentified certain aspects from long review texts, thus, fine-tuning on larger datasets in the domain is highly recommended.
4. Our models currently identified 2 polarities i.e positive and negative. Further fine-tuning should be added to conducted to address neutral sentiments

These limitations underscore the need for future research to address these gaps and enhance the robustness of our findings.

6 Conclusion

In summary, this study performed ABSA by combining AOE with semantic search and SA techniques to extract nuanced insights into consumers' specific aspect preferences from smartphone reviews from prominent brands on Amazon. The model performance was then assessed through both manual evaluation and Random Forest prediction model. The outputs were then analysed to offer actionable recommendations for product development like prioritising battery performance, marketing strategies, and competitive positioning based on consumer sentiments and brand performance. We also acknowledge several limitations,

including language bias, limited attribute focus, and difficulties on dealing with neutral sentiment, which can be solved by further fine-tuning procedures.

Appendices

Appendix 1 – Manual Evaluation Result

Review	Aspect	Opinion	SA Score	Manual SA	Status
The camera / picture quality is superb and overall it's fast, slick and the battery life is excellent	camera	superb	0.6249	positive	aligned
The camera / picture quality is superb and overall it's fast, slick and the battery life is excellent	battery	excellent	0.5719	positive	aligned
The device in its new titanium finish is superb	design	superb	0.6249	positive	aligned
Bought this phone refurbished. Never again. Too many issues. Terrible battery retention. Issues with touch screen.	display	terrible	-0.4767	negative	Partially aligned
Love the picture quality. Definately great value for money. I also like the curved appearance of the screen. Definately recommend.	camera	love	0.5	positive	aligned
Love the picture quality. Definately great value for money. I also like the curved appearance of the screen. Definately recommend.	display	good	0.5	positive	aligned
Nice phone with great camera quality and good battery quality	camera	great	0.8	positive	aligned
Nice phone with great camera quality and good battery quality	battery	good	0.7	positive	aligned
Works great and really good camera . No faults at all	camera	good	0.4404	positive	aligned

Appendix 2 – Chat GPT prompts

ChatGPT 3.5 ▾

 You
from pyabsa import AspectSentimentTripletExtraction as ASTE

Load the model
triplet_extractor = ASTE.AspectSentimentTripletExtractor("english")
triplet_extractor = ASTE.AspectSentimentTripletExtractor("multilingual")

Predict
examples = [
 "I would like to have volume buttons rather than the adjustment that is on the front .#####[([5, 6], [2], 'NEG')]",
 "It runs perfectly .#####[([1], [2], 'POS')]",
 "Sometimes the screen even goes black on this computer .#####[([2], [5], 'NEG')]",
 "Its fast and another thing I like is that it has three USB ports .#####[([12, 13], [6], 'POS')]",
]
for example in examples:
 triplet_extractor.predict(example)

 ChatGPT
It seems like you're using the PyABSA library for Aspect-Based Sentiment Analysis (ABSA). However, there might be a small issue in your code. The `predict` method doesn't return anything. Instead, it directly prints the results. To capture the results or manipulate them further, you should consider modifying your code to store the results in a variable. Here's how you can do it:

```
python
from pyabsa import AspectSentimentTripletExtraction as ASTE
```

Message ChatGPT

ChatGPT 3.5 ▾

 You

```
if __name__ == "__main__":
    generator = ABSAInstruction.ABSAGenerator("multilingual")
    example = [
        "The food is good, but the service is bad.",
        "The laptop is good, but the battery life is too short and the camera is blurry.",
    ]

    for example in example:
        result = generator.predict(example)
        print(result)
```

 ChatGPT

Your code snippet seems to be using the ABSAGenerator from the PyABSA library to perform Aspect-Based Sentiment Analysis (ABSA) on multilingual text. The `predict` method of the generator will provide the aspect-based sentiment analysis results for each example.

Here's your code with a minor correction:

```
python
Copy code
if __name__ == "__main__":
    from pyabsa import ABSAInstruction
    generator = ABSAInstruction.ABSAGenerator("multilingual")
```

 Message ChatGPT

ChatGPT 3.5 ▾

```
if __name__ == "__main__":
    from pyabsa import ABSAInstruction

    generator = ABSAInstruction.ABSAGenerator("multilingual")
    examples = [
        "The food is good, but the service is bad.",
        "The laptop is good, but the battery life is too short and the camera is blur
    ]

    for example in examples:
        result = generator.predict(example)
        print(result)
```

This will iterate through each example, perform ABSA on it using the multilingual model, and print the result. Make sure to import `ABSAInstruction` at the beginning of your script.



Ha You

what if my example now is a column in the dataframe

ChatGPT

If your examples are stored in a DataFrame column and you want to perform Aspect-Based Sentiment Analysis (ABSA) on each example in that column, you can iterate through the DataFrame and apply the ABSA generator to each example. Here's how you can do it:



python

Copy code

Message ChatGPT



**You**

what if my example now is a column in the dataframe

**ChatGPT**

If your examples are stored in a DataFrame column and you want to perform Aspect-Based Sentiment Analysis (ABSA) on each example in that column, you can iterate through the DataFrame and apply the ABSA generator to each example. Here's how you can do it:

```
python Copy code

import pandas as pd
from pyabsa import ABSAInstruction

if __name__ == "__main__":
    # Assuming you have a DataFrame df with a column 'example'
    df = pd.DataFrame({'example': [
        "The food is good, but the service is bad.",
        "The laptop is good, but the battery life is too short and the camera quality is poor."
    ]})

    # Initialize the ABSA generator
    generator = ABSAInstruction.ABSAGenerator("multilingual")

    # Apply ABSA on each example in the DataFrame column
    for example in df['example']:
        result = generator.predict(example)
        print(result)
```

This code will iterate through each example in the 'example' column of your DataFrame, apply ABSA using the multilingual model, and print the result for each example. Make sure to adjust the DataFrame variable name (`df`) and column name ('example') according to your actual DataFrame structure.

**You**

df['review_text'] here's the column

**ChatGPT**

If `df['review_text']` contains the text data on which you want to perform Aspect-Based Sentiment Analysis (ABSA), you can use the following code snippet to iterate through the column and apply the ABSA generator to each text:

```
python Copy code

import pandas as pd
from pyabsa import ABSAInstruction

if __name__ == "__main__":
    # Assuming you have a DataFrame df with a column 'review_text'
    df = pd.DataFrame({'review_text': [
```

Message ChatGPT ↑

ChatGPT can make mistakes. Check important info.

 You
df['review_text'] here's the column

 ChatGPT
If `df['review_text']` contains the text data on which you want to perform Aspect-Based Sentiment Analysis (ABSA), you can use the following code snippet to iterate through the column and apply the ABSA generator to each text:

```
python Copy code

import pandas as pd
from pyabsa import ABSAInstruction

if __name__ == "__main__":
    # Assuming you have a DataFrame df with a column 'review_text'
    df = pd.DataFrame({'review_text': [
        "The food is good, but the service is bad.",
        "The laptop is good, but the battery life is too short and the camera quality is poor."
    ]})

    # Initialize the ABSA generator
    generator = ABSAInstruction.ABSAGenerator("multilingual")

    # Apply ABSA on each review text in the DataFrame column
    for review_text in df['review_text']:
        result = generator.predict(review_text)
        print(result)
```

This script will iterate through each review text in the 'review_text' column of your DataFrame, apply ABSA using the multilingual model, and print the result for each review text. Make sure to adjust the DataFrame variable name (``df``) and column name ('review_text') according to your actual DataFrame structure.

 You
no you dont understand, give me the way to replace example with df['review_text']

 ChatGPT
Got it, you want to replace the example variable with the contents of the 'review_text' column in your DataFrame. Here's how you can modify the code to achieve that:

```
python Copy code

import pandas as pd
from pyabsa import ABSAInstruction

if __name__ == "__main__":
    # Assuming you have a DataFrame df with a column 'review_text'
    df = pd.DataFrame({'review_text': [
        "The food is good, but the service is bad."
    ]})
```

Message ChatGPT ↑

ChatGPT can make mistakes. Check important info.

Ha

You

save the result into a column in df

ChatGPT

To save the ABSA results into a new column in your DataFrame `df`, you can create an empty list to store the results and then assign that list to a new column in the DataFrame. Here's how you can do it:

```
python Copy code

import pandas as pd
from pyabsa import ABSAInstruction

if __name__ == "__main__":
    # Assuming you have a DataFrame df with a column 'review_text'
    df = pd.DataFrame({'review_text': [
        "The food is good, but the service is bad.",
        "The laptop is good, but the battery life is too short and the camera is not great."
    ]})

    # Initialize the ABSA generator
    generator = ABSAInstruction.ABSAGenerator("multilingual")

    # Create an empty list to store the ABSA results
    absa_results = []

    # Apply ABSA on each review text in the DataFrame column
    for review_text in df['review_text']:
        result = generator.predict(review_text)
        absa_results.append(result)

    # Add a new column 'absa_result' to the DataFrame and assign the ABSA results
    df['absa_result'] = absa_results

    # Print the updated DataFrame
    print(df)
```

This code will iterate through each review text in the 'review_text' column of your DataFrame, apply ABSA using the multilingual model, store the results in the `absa_results` list, add a new column 'absa_result' to the DataFrame, and finally print the updated DataFrame with the ABSA results included. Adjust the column names and DataFrame structure as per your actual data.

Ha

You

{'text': 'looks brand new, works amazing, only issue is the battery life, not too big of a deal, but i was expecting a little above 89% would highly recommend if you want a decent phone on a budget!\n', 'Quadruples': [{aspect: 'NULL', 'polarity': 'positive', 'opinion': 'new', 'category': 'LAPTOP#DESIGN_FEATURES'}, {aspect: 'battery life', 'polarity': 'negative', 'opinion': 'amazing', 'category': 'LAPTOP#OPERATION_PERFORMANCE'}]} here's my output saved in df['absa_result']. Give me the code to extract 3 columns: aspect, polarity and opinion

Message ChatGPT ↑

ChatGPT can make mistakes. Check important info.

Ha**You**

{'text': 'looks brand new, works amazing, only issue is the battery life, not too big of a deal, but i was expecting a little above 89% would highly recommend if you want a decent phone on a budget!\n', 'Quadruples': [{aspect: 'NULL', 'polarity': 'positive', 'opinion': 'new', 'category': 'LAPTOP#DESIGN_FEATURES'}, {aspect: 'battery life', 'polarity': 'negative', 'opinion': 'amazing', 'category': 'LAPTOP#OPERATION_PERFORMANCE'}]} here's my output saved in df['absa_result']. Give me the code to extract 3 columns: aspect, polarity and opinion

**ChatGPT**

You can extract the 'aspect', 'polarity', and 'opinion' from the 'absa_result' column in your DataFrame and create new columns for each of them. Here's how you can do it:

```
python
Copy code

import pandas as pd

# Assuming df['absa_result'] contains the ABSA results in the specified format
data = {'text': ['looks brand new, works amazing, only issue is the battery life, not too big of a deal, but i was expecting a little above 89% would highly recommend if you want a decent phone on a budget!'],
        'absa_result': [{}{'text': 'looks brand new, works amazing, only issue is the battery life, not too big of a deal, but i was expecting a little above 89% would highly recommend if you want a decent phone on a budget!', 'Quadruples': [{aspect: 'NULL', 'polarity': 'positive', 'opinion': 'new', 'category': 'LAPTOP#DESIGN_FEATURES'}, {aspect: 'battery life', 'polarity': 'negative', 'opinion': 'amazing', 'category': 'LAPTOP#OPERATION_PERFORMANCE'}]}]}

df = pd.DataFrame(data)

# Extract 'aspect', 'polarity', and 'opinion' from the 'absa_result' column
df['aspect'] = df['absa_result'].apply(lambda x: [item['aspect'] for item in x['Quadruples']])
df['polarity'] = df['absa_result'].apply(lambda x: [item['polarity'] for item in x['Quadruples']])
df['opinion'] = df['absa_result'].apply(lambda x: [item['opinion'] for item in x['Quadruples']])

# If you want to flatten the lists to single values, you can do so by accessing the first element
df['aspect'] = df['aspect'].apply(lambda x: x[0] if x else None)
df['polarity'] = df['polarity'].apply(lambda x: x[0] if x else None)
df['opinion'] = df['opinion'].apply(lambda x: x[0] if x else None)

# Print the DataFrame with the extracted columns
print(df[['aspect', 'polarity', 'opinion']])
```

This code will create three new columns ('aspect', 'polarity', and 'opinion') in your DataFrame `df` containing the extracted values from the 'absa_result' column. Adjust the column names and DataFrame structure as per your actual data.

Ha**You**

```
# Set display options
pd.set_option('display.max_colwidth', None) # Ensure no truncation
pd.set_option('display.max_rows', None) # Display any number of rows

for query in defined_aspects:
    top_sentences = semantic_search(query, np.array(list(df['sbert_embedding'])))
    for sentence in top_sentences:
        print(sentence)
```

Message ChatGPT



ChatGPT can make mistakes. Check important info.

 You

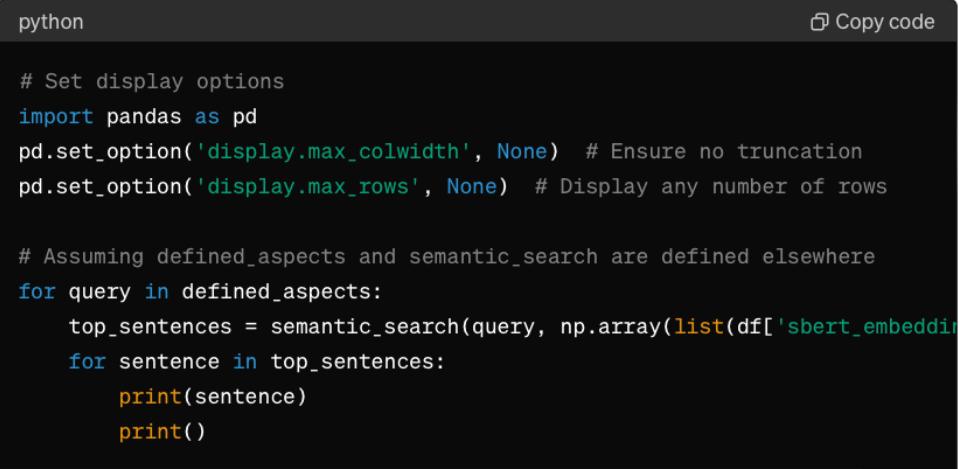
```
# Set display options
pd.set_option('display.max_colwidth', None) # Ensure no truncation
pd.set_option('display.max_rows', None) # Display any number of rows

for query in defined_aspects:
    top_sentences = semantic_search(query, np.array(list(df['sbert_embedding'])))
for sentence in top_sentences:
    print(sentence)
    print()
```

 ChatGPT

It seems like you're setting up your display options and then iterating through defined aspects to perform semantic search on the sentences in your DataFrame. However, the code snippet you provided is incomplete, as the `semantic_search` function and the `defined_aspects` variable are not defined here.

Assuming you have these defined elsewhere, you can integrate them into your loop like this:



```
python
Copy code

# Set display options
import pandas as pd
pd.set_option('display.max_colwidth', None) # Ensure no truncation
pd.set_option('display.max_rows', None) # Display any number of rows

# Assuming defined_aspects and semantic_search are defined elsewhere
for query in defined_aspects:
    top_sentences = semantic_search(query, np.array(list(df['sbert_embedding'])))
    for sentence in top_sentences:
        print(sentence)
        print()
```

Make sure to replace `semantic_search(query, np.array(list(df['sbert_embedding'])))` with the correct call to your semantic search function. Also, ensure that `defined_aspects` is properly defined with the aspects you want to search for.



 You

what's in my top_sentences?

 ChatGPT

To determine what's in the `top_sentences`, you need to inspect its content. Since you're iterating over `top_sentences` and printing each sentence, it seems like `top_sentences` is a list of sentences.

Here's how you can print the content of `top_sentences` to see what it contains:



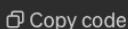
Message ChatGPT

ChatGPT can make mistakes. Check important info.

 **You**
what's in my top_sentences?

 **ChatGPT**
To determine what's in the `top_sentences`, you need to inspect its content. Since you're iterating over `top_sentences` and printing each sentence, it seems like `top_sentences` is a list of sentences.

Here's how you can print the content of `top_sentences` to see what it contains:

```
python 
```

```
for sentence in top_sentences:
    print(sentence)
```

By running this code, you'll print each sentence in `top_sentences`, allowing you to see its content and structure. This will give you a clear view of what's contained in `top_sentences`.

 **You**
from sklearn.metrics.pairwise import cosine_similarity

```
def semantic_search(query, embeddings, top_k=1):
    query_embedding = sbert_model.encode([query])
    cos_similarities = cosine_similarity(query_embedding, embeddings)[0]
    top_k_indices = np.argsort(cos_similarities)[-1:][:-top_k]
    return df.iloc[top_k_indices]['aspect']# Set display options
pd.set_option('display.max_colwidth', None) # Ensure no truncation
pd.set_option('display.max_rows', None) # Display any number of rows

for query in defined_aspects:
    defined_aspects = semantic_search(query, np.array(list(df['sbert_embedding'])))
```

 **ChatGPT**
In your `semantic_search` function, you're calculating cosine similarities between a query embedding and a matrix of embeddings. Then, you're retrieving the top-k indices of the most similar embeddings and returning the corresponding 'aspect' from your DataFrame.

Now, since you're iterating over `defined_aspects` (which is used as a query), and calling the `semantic_search` function for each query, `semantic_search` will return the 'aspect' corresponding to the most similar sentence embedding to each query.

In your loop, you're storing the result of `semantic_search` in `defined_aspects`, which might overwrite your original `defined_aspects` list. You may want to rename one of them to avoid confusion.

To see what's in `defined_aspects` after calling `semantic_search`, you can print it:

```
python 
```

```
for query in defined_aspects:
    top_aspects = semantic_search(query, np.array(list(df['sbert_embedding'])))
    print(top_aspects)
```

Message ChatGPT 

ChatGPT can make mistakes. Check important info.

Ha

You
what if I want to save the defined_aspects to a column in df

**ChatGPT**

To save the defined aspects to a new column in your DataFrame `df`, you can modify your loop to store the aspects in a list and then assign that list to a new column in the DataFrame. Here's how you can do it:

```
python Copy code

import pandas as pd
from sklearn.metrics.pairwise import cosine_similarity

def semantic_search(query, embeddings, top_k=1):
    query_embedding = sbert_model.encode([query])
    cos_similarities = cosine_similarity(query_embedding, embeddings)[0]
    top_k_indices = np.argsort(cos_similarities)[::-1][:top_k]
    return df.iloc[top_k_indices]['aspect']

# Sample DataFrame
df = pd.DataFrame({'sbert_embedding': [], 'defined_aspects': []})

# Sample defined_aspects
defined_aspects = ['price', 'design', 'performance']

# Set display options
pd.set_option('display.max_colwidth', None) # Ensure no truncation
pd.set_option('display.max_rows', None) # Display any number of rows

# Iterate through defined_aspects and save the results to a new column in df
aspect_results = []
for query in defined_aspects:
    top_aspects = semantic_search(query, np.array(list(df['sbert_embedding'])))
    aspect_results.append(top_aspects)

# Add a new column 'defined_aspects' to the DataFrame and assign the aspect results
df['defined_aspects'] = aspect_results

# Print the updated DataFrame
print(df)
```

This code will iterate through each aspect in `defined_aspects`, perform semantic search, store the results in a list, add a new column 'defined_aspects' to the DataFrame, and finally print the updated DataFrame with the defined aspects included. Adjust the code according to your actual data and requirements.

**Ha**

You
if __name__ == "__main__":
 # Initialize the ABSA generator



Message ChatGPT



You

```

if __name__ == "__main__":
    # Initialize the ABSA generator
    generator = ABSAInstruction.ABSAGenerator("multilingual")

    # Create an empty list to store the ABSA results
    absa_results = []

    # Apply ABSA on each review text in the DataFrame column
    for review_text in df['review_text'].loc[10]:
        result = generator.predict(review_text)
        absa_results.append(result)

    # Add a new column 'absa_result' to the DataFrame and assign the ABSA results
    df['ABSA_result'] = absa_results

```



ChatGPT

It looks like you're initializing an ABSA generator, applying it to each review text in a DataFrame column, and storing the results in a new column named 'ABSA_result'. However, the `for` loop appears to be iterating over a single review text (`df['review_text'].loc[10]`), which may not be what you intend.

If you want to apply ABSA to every review text in the 'review_text' column, you should iterate over the entire column rather than just a single entry. Here's how you can modify your code:

```

python                                         ⌂ Copy code

if __name__ == "__main__":
    import pandas as pd
    from pyabsa import ABSAInstruction

    # Assuming you have a DataFrame df with a column 'review_text'
    df = pd.DataFrame({'review_text': [
        "The food is good, but the service is bad.",
        "The laptop is good, but the battery life is too short and the camera is not good."
    ]})

    # Initialize the ABSA generator
    generator = ABSAInstruction.ABSAGenerator("multilingual")

    # Create an empty list to store the ABSA results
    absa_results = []

    # Apply ABSA on each review text in the DataFrame column
    for review_text in df['review_text']:
        result = generator.predict(review_text)
        absa_results.append(result)

    # Add a new column 'ABSA_result' to the DataFrame and assign the ABSA results
    df['ABSA_result'] = absa_results

```

Message ChatGPT



ChatGPT can make mistakes. Check important info.

This code will apply ABSA to each review text in the 'review_text' column, store the results in a new column named 'ABSA_result', and then print the updated DataFrame. Adjust the code according to your actual data and requirements.

**You**

i just want the first 10 rows from df['review_text']

**ChatGPT**

If you want to apply ABSA to only the first 10 rows of the 'review_text' column in your DataFrame `df`, you can use DataFrame slicing to select those rows before iterating over them. Here's how you can do it:

```
python Copy code
if __name__ == "__main__":
    import pandas as pd
    from pyabsa import ABSAInstruction

    # Assuming you have a DataFrame df with a column 'review_text'
    df = pd.DataFrame({'review_text': [
        "The food is good, but the service is bad.",
        "The laptop is good, but the battery life is too short and the camera quality is poor."
    ]})

    # Initialize the ABSA generator
    generator = ABSAInstruction.ABSAGenerator("multilingual")

    # Create an empty list to store the ABSA results
    absa_results = []

    # Apply ABSA on each review text in the first 10 rows of the DataFrame
    for review_text in df['review_text'].iloc[:10]: # Selecting only the first 10 rows
        result = generator.predict(review_text)
        absa_results.append(result)

    # Add a new column 'ABSA_result' to the DataFrame and assign the ABSA results
    df['ABSA_result'] = absa_results

    # Print the updated DataFrame
    print(df)
```

This code will apply ABSA to each review text in the first 10 rows of the 'review_text' column, store the results in a new column named 'ABSA_result', and then print the updated DataFrame. Adjust the code according to your actual data and requirements.

**You**

aspect \
 0 [iPhone 13 Pro Max]



Message ChatGPT



ChatGPT can make mistakes. Check important info.

Ha

You

```

aspect \
0 [iPhone 13 Pro Max]
1 [device, camera / picture quality, battery lif...
2 [NULL]
3 [media, NULL]
4 [phone]
5 [iPhone]
6 [phone]
7 [phone]
8 [camera, Battery life]
9 [phone]

polarity \
0 [positive]
1 [positive, positive, positive, positive]
2 [positive]
3 [negative, positive]
4 [negative]
5 [positive]
6 [positive]
7 [positive]
8 [positive, positive]
9 [positive]

opinion \
0 [great]
1 [superb, superb, excellent, worth]
2 [loves]
3 [NULL, NULL]
4 [NULL]
5 [love]
6 [Brilliant]
7 [reorder]
8 [better, better]
9 [love]

category
0 [LAPTOP#OPERATION_PERFORMANCE]
1 [LAPTOP#DESIGN_FEATURES, DISPLAY#QUALITY, BATT...
2 [LAPTOP#GENERAL]
3 [MULTIMEDIA_DEVICES#OPERATION_PERFORMANCE, MUL...
4 [LAPTOP#QUALITY]
5 [LAPTOP#GENERAL]
6 [LAPTOP#GENERAL]
7 [LAPTOP#GENERAL]
8 [LAPTOP#GENERAL, BATTERY#GENERAL]
9 [LAPTOP#GENERAL] some of my aspects have multiple values, split to
different rows, keeping the same values for other column

```

ChatGPT

It looks like you have aspects, polarities, opinions, and categories stored as lists within your DataFrame, with some aspects having multiple values. You want to split these lists into different rows while keeping the same values for other columns.



You can achieve this by using the `explode` function in pandas to split the lists into separate rows.

Message ChatGPT



ChatGPT can make mistakes. Check important info.

 ChatGPT

It looks like you have aspects, polarities, opinions, and categories stored as lists within your DataFrame, with some aspects having multiple values. You want to split these lists into different rows while keeping the same values for other columns.

You can achieve this by using the `explode` function in pandas to split the lists into separate rows. Here's how you can do it:

```
python Copy code

import pandas as pd

# Sample DataFrame
data = {
    'aspect': [
        ['iPhone 13 Pro Max'],
        ['device', 'camera / picture quality', 'battery life', 'operating system'],
        ['NULL'],
        ['media', 'NULL'],
        ['phone'],
        ['iPhone'],
        ['phone'],
        ['phone'],
        ['camera', 'Battery life'],
        ['phone']
    ],
    'polarity': [
        ['positive'],
        ['positive', 'positive', 'positive', 'positive'],
        ['positive'],
        ['negative', 'positive'],
        ['negative'],
        ['positive'],
        ['positive'],
        ['positive'],
        ['positive', 'positive'],
        ['positive']
    ],
    'opinion': [
        ['great'],
        ['superb', 'superb', 'excellent', 'worth'],
        ['loves'],
        ['NULL', 'NULL'],
        ['NULL'],
        ['love'],
        ['Brilliant'],
        ['reorder'],
        ['better', 'better'],
        ['love']
    ],
    'category': [

```

Message ChatGPT ↑

ChatGPT can make mistakes. Check important info.

This will split the lists in the 'aspect', 'polarity', 'opinion', and 'category' columns into separate rows while duplicating the other values in the DataFrame accordingly. Adjust the code according to your actual DataFrame structure and requirements.

**You**

how can I remove "LAPTOP" or "#", from each row in my category column

**ChatGPT**

To remove "LAPTOP" or "#" from each row in your category column, you can use the ``str.replace()`` method in pandas. Here's how you can do it:

```
python Copy code
import pandas as pd

# Sample DataFrame
data = {
    'category': [
        ['LAPTOP'],
        ['camera / picture quality', 'battery life', 'operating system'],
        ['NULL'],
        ['media', 'NULL'],
        ['phone'],
        ['iPhone'],
        ['phone'],
        ['phone'],
        ['camera', 'Battery life'],
        ['phone']
    ],
    'polarity': [
        ['positive'],
        ['positive', 'positive', 'positive', 'positive'],
        ['positive'],
        ['negative', 'positive'],
        ['negative'],
        ['positive'],
        ['positive'],
        ['positive'],
        ['positive', 'positive'],
        ['positive']
    ],
    'opinion': [
        ['great'],
        ['superb', 'superb', 'excellent', 'worth'],
        ['loves'],
        ['NULL', 'NULL'],
        ['NULL'],
        ['love'],
        ['Brilliant'],
        ['cool']
    ]
}
```

Message ChatGPT ↑

ChatGPT can make mistakes. Check important info.

 You

but it split my string to seperate letters?

 ChatGPT

I see, if the strings are being split into individual letters, it suggests that each category entry is being treated as a list of strings rather than a single string. To remove "LAPTOP" or "#" from each entry in the 'category' column while keeping the entries as strings, you can use the `str.replace()` method directly on the column. Here's how you can do it:

```
python  Copy code
```

```
import pandas as pd

# Sample DataFrame
data = {
    'aspect': [
        'iPhone 13 Pro Max',
        'device, camera / picture quality, battery life, operating system',
        'NULL',
        'media, NULL',
        'phone',
        'iPhone',
        'phone',
        'phone',
        'camera, Battery life',
        'phone'
    ],
    'polarity': [
        'positive',
        'positive, positive, positive, positive',
        'positive',
        'negative, positive',
        'negative',
        'positive',
        'positive',
        'positive',
        'positive, positive',
        'positive'
    ],
    'opinion': [
        'great',
        'superb, superb, excellent, worth',
        'loves',
        'NULL, NULL',
        'NULL',
        'love',
        'Brilliant',
        'reorder',
        'better, better',
        'love'
    ]
}
```


Message ChatGPT 

ChatGPT can make mistakes. Check important info.

This code directly applies the `str.replace()` method to the 'category' column of the DataFrame, removing "LAPTOP" and "#" from each entry in the column. Adjust the code according to your actual DataFrame structure and requirements.

**You**

only the first letter is kept

**ChatGPT**

I see, it seems like the entries in the 'category' column are being treated as individual strings rather than a single string. To avoid this issue, you can first convert each entry in the 'category' column to a single string before applying the `str.replace()` method. Here's how you can do it:

```
python Copy code
import pandas as pd

# Sample DataFrame
data = {
    'aspect': [
        'iPhone 13 Pro Max',
        'device, camera / picture quality, battery life, operating system',
        'NULL',
        'media, NULL',
        'phone',
        'iPhone',
        'phone',
        'phone',
        'camera, Battery life',
        'phone'
    ],
    'polarity': [
        'positive',
        'positive, positive, positive, positive',
        'positive',
        'negative, positive',
        'negative',
        'positive',
        'positive',
        'positive',
        'positive, positive',
        'positive'
    ],
    'opinion': [
        'great',
        'superb, superb, excellent, worth',
        'loves',
        'NULL, NULL',
        'NULL',
        'love'
    ]
}
```

Message ChatGPT



ChatGPT can make mistakes. Check important info.

You
only the first letter is kept

ChatGPT

I see, it seems like the `str.replace()` method is not behaving as expected, and only the first letter is being kept after the replacement. To resolve this, we can use regular expressions (regex) to perform more specific replacements. Here's how you can do it:

```
python
import pandas as pd

# Sample DataFrame
data = {
    'aspect': [
        'iPhone 13 Pro Max',
        'device, camera / picture quality, battery life, operating system',
        'NULL',
        'media, NULL',
        'phone',
        'iPhone',
        'phone',
        'phone',
        'camera, Battery life',
        'phone'
    ],
    'polarity': [
        'positive',
        'positive, positive, positive, positive',
        'positive',
        'negative, positive',
        'negative',
        'positive',
        'positive',
        'positive',
        'positive, positive',
        'positive'
    ],
    'opinion': [
        'great',
        'superb, superb, excellent, worth',
        'loves',
        'NULL, NULL',
        'NULL',
        'love',
        'Brilliant',
        'reorder',
        'better, better',
        'love'
    ]
}
```

Message ChatGPT



ChatGPT can make mistakes. Check important info.

You

what's the difference from the previous code??????

ChatGPT

My apologies for the confusion. It seems I made an oversight. Both the previous and the current code use the same regex pattern ` \bLAPTOP\b | #`, which should match either the whole word "LAPTOP" or the "#" symbol.

The issue might lie elsewhere in your code or data. Let's try a different approach. We can split the 'category' column by "#" and then join the resulting parts except for the first one. Here's the updated approach:

```
python  Copy code

import pandas as pd

# Sample DataFrame
data = {
    'aspect': [
        'iPhone 13 Pro Max',
        'device, camera / picture quality, battery life, operating system',
        'NULL',
        'media, NULL',
        'phone',
        'iPhone',
        'phone',
        'phone',
        'camera, Battery life',
        'phone'
    ],
    'polarity': [
        'positive',
        'positive, positive, positive, positive',
        'positive',
        'negative, positive',
        'negative',
        'positive',
        'positive',
        'positive',
        'positive, positive',
        'positive'
    ],
    'opinion': [
        'great',
        'superb, superb, excellent, worth',
        'loves',
        'NULL, NULL',
        'NULL',
        'love',
        'Brilliant',
        ' '
    ]
}
```

Message ChatGPT

ChatGPT can make mistakes. Check important info.

 You
it's blank

 ChatGPT

If the resulting 'category' column is blank, it indicates that the split operation did not occur as expected, or there might be some issue with the data. To troubleshoot, let's print the intermediate steps to understand what's happening:

```
python
import pandas as pd

# Sample DataFrame
data = {
    'aspect': [
        'iPhone 13 Pro Max',
        'device, camera / picture quality, battery life, operating system',
        'NULL',
        'media, NULL',
        'phone',
        'iPhone',
        'phone',
        'phone',
        'camera, Battery life',
        'phone'
    ],
    'polarity': [
        'positive',
        'positive, positive, positive, positive',
        'positive',
        'negative, positive',
        'negative',
        'positive',
        'positive',
        'positive',
        'positive, positive',
        'positive'
    ],
    'opinion': [
        'great',
        'superb, superb, excellent, worth',
        'loves',
        'NULL, NULL',
        'NULL',
        'love',
        'Brilliant',
        'reorder',
        'better, better',
        'love'
    ],
    'category': []
}
```

Message ChatGPT



ChatGPT can make mistakes. Check important info.

 You

```
usr/local/lib/python3.10/dist-packages/ipykernel/ipkernel.py:283: DeprecationWarning:
`should_run_async` will not call `transform_cell` automatically in the future. Please pass the
result to `transformed_cell` argument and any exception that happen during the transform in
`preprocessing_exc_tuple` in IPython 7.17 and above.
and should_run_async(code)
ERROR: unknown command "sentence_transformers"
<frozen importlib._bootstrap>:914: ImportWarning: _PyDrive2ImportHook.find_spec() not
found; falling back to find_module()
<frozen importlib._bootstrap>:914: ImportWarning: _PyDriveImportHook.find_spec() not
found; falling back to find_module()
<frozen importlib._bootstrap>:914: ImportWarning: _GenerativeAIImportHook.find_spec() not
found; falling back to find_module()
<frozen importlib._bootstrap>:914: ImportWarning: _OpenCVImportHook.find_spec() not
found; falling back to find_module()
<frozen importlib._bootstrap>:914: ImportWarning: APICoreClientInfoImportHook.find_spec() not
found; falling back to find_module()
<frozen importlib._bootstrap>:914: ImportWarning: _BokehImportHook.find_spec() not found;
falling back to find_module()
<frozen importlib._bootstrap>:914: ImportWarning: _AltairImportHook.find_spec() not found;
falling back to find_module()

-----
ModuleNotFoundError          Traceback (most recent call last)
<ipython-input-10-82d747d77dfa> in <cell line: 2>()
      1 get_ipython().system('pip sentence_transformers')
----> 2 from sentence_transformers import SentenceTransformer
      3
      4 # Load the model
      5 sbert_model = SentenceTransformer('all-MiniLM-L6-v2')

ModuleNotFoundError: No module named 'sentence_transformers'
```



ChatGPT

It seems there was an error related to importing the `'sentence_transformers'` module. It appears that you might have tried to install the package using the command `'!pip sentence_transformers'`, which is incorrect.

To install the `'sentence_transformers'` package, you should use the following command:

```
diff
Copy code
!pip install sentence-transformers
```

Once you've installed the package correctly, you can import `'SentenceTransformer'` without any issues.

If you have already installed the package but are still encountering import errors, it might be due to an environment issue or the package not being installed in the correct Python environment. Ensure that you're working in the correct environment and that the package is installed properly.

 You

`!pip install sentence-transformers`



Message ChatGPT



ChatGPT can make mistakes. Check important info.

 You

```

!pip install sentence-transformers
from sentence_transformers import SentenceTransformer

# Load the model
sbert_model = SentenceTransformer('all-MiniLM-L6-v2')

# Create embeddings
test_df['sbert_embedding'] = [sbert_model.encode(text) for text in test_df['category']]
from sklearn.metrics.pairwise import cosine_similarity

def semantic_search(query, embeddings, top_k=1):
    query_embedding = sbert_model.encode([query])
    cos_similarities = cosine_similarity(query_embedding, embeddings)[0]
    top_k_indices = np.argsort(cos_similarities)[-1:-1:-1][top_k]
    return test_df.iloc[top_k_indices]['category']

# Define aspects
defined_aspects = ['camera','battery','display','processor','storage','price','design']# Set
display options
pd.set_option('display.max_colwidth', None) # Ensure no truncation
pd.set_option('display.max_rows', None) # Display any number of rows

# Iterate through defined_aspects and save the results to a new column in df
aspect_results = []
for category in test_df['category']:
    top_aspects = semantic_search(category, np.array(list(test_df['sbert_embedding'])))
    aspect_results.append(top_aspects)

# Add a new column 'defined_aspects' to the DataFrame and assign the aspect results
test_df['defined_aspects'] = aspect_results

```



ChatGPT

It seems you're trying to use the '**SentenceTransformer**' to embed text data and then perform semantic search to find similar categories based on embeddings. However, I noticed a few issues:

1. You're trying to create embeddings for the 'category' column, but the '**SentenceTransformer**' model is initialized with the argument 'all-MiniLM-L6-v2'. This model is designed for generating sentence embeddings, but it might not be the most suitable for representing categories. You might want to consider using a different model or approach for embedding categories.
2. The '**semantic_search**' function is currently returning the entire row from the DataFrame instead of the aspect. It seems you intended to return the category instead of the entire row.
3. You're using the 'category' column both as input to semantic search and as the output. This might not be the intended behavior. Usually, semantic search takes a query and searches for similar items in a corpus. It's unclear how you're intending to use the category information in this context.

Could you please provide more details on what you're trying to achieve with this code so I can assist you further? Specifically, what is the goal of semantic search in your application, and how do you intend to use the results?



You

I have a list of defined_aspects, for each value in test_df['category'], I want to find the closest meaning word from that defined_aspects and return the word from defined_aspects and save to a separate column in test_df

Message ChatGPT



ChatGPT can make mistakes. Check important info.



You

I have a list of defined_aspects, for each value in test_df['category'], I want to find the closest meaning word from that defined_aspects and return the word from defined_aspects and save to a seperate column in test_df



ChatGPT

Got it! To achieve your goal, we need to find the closest defined aspect for each category in the `test_df['category']` column. We can use the semantic similarity between the embeddings of categories and defined aspects to accomplish this.

Here's the updated code:

```
python Copy code
from sentence_transformers import SentenceTransformer
import pandas as pd
from sklearn.metrics.pairwise import cosine_similarity

# Load the SentenceTransformer model
sbert_model = SentenceTransformer('all-MiniLM-L6-v2')

# Example DataFrame
test_df = pd.DataFrame({'category': ['camera quality', 'battery life', 'processor power', 'display resolution', 'storage capacity'], 'defined_aspect': ['camera', 'battery', 'display', 'processor', 'storage']})

# Define aspects
defined_aspects = ['camera', 'battery', 'display', 'processor', 'storage']

# Function to find the closest aspect for a given category
def find_closest_aspect(category, defined_aspects):
    category_embedding = sbert_model.encode([category])
    aspect_embeddings = sbert_model.encode(defined_aspects)
    similarities = cosine_similarity(category_embedding, aspect_embeddings)
    closest_aspect_index = similarities.argmax()
    closest_aspect = defined_aspects[closest_aspect_index]
    return closest_aspect

# Apply the function to each category in test_df and save the result to a new column
test_df['closest_aspect'] = test_df['category'].apply(lambda x: find_closest_aspect(x, defined_aspects))

# Display the updated DataFrame
print(test_df)
```

This code defines a function `find_closest_aspect` that takes a category and a list of defined aspects, computes the semantic similarity between the category and each aspect using cosine similarity, and returns the closest aspect. Then, it applies this function to each category in the `test_df['category']` column and saves the result to a new column called `closest_aspect`.

Adjust the code according to your actual DataFrame structure and requirements. Let me know if you need further assistance!



Message ChatGPT



ChatGPT can make mistakes. Check important info.

 **You**

for opinion in test_df['opinion']:
 sentence_sentiment = get_sentiment(opinion) how can I save this to test_df as a column

 **ChatGPT**

To save the sentiment of each opinion in the 'opinion' column of `test_df` as a new column, you can loop through each opinion, get its sentiment using a function (assuming `get_sentiment` is a function you have defined elsewhere), and then save the sentiment values to a new column in `test_df`. Here's how you can do it:

python

 Copy code

```
# Loop through each opinion in test_df['opinion']
sentiments = []
for opinion in test_df['opinion']:
    # Get sentiment for each opinion using the get_sentiment function
    sentence_sentiment = get_sentiment(opinion)
    sentiments.append(sentence_sentiment)

# Add the list of sentiments as a new column 'opinion_sentiment' to test_df
test_df['opinion_sentiment'] = sentiments
```

This code iterates through each opinion in the 'opinion' column, retrieves its sentiment using the `get_sentiment` function, and appends the sentiment values to a list called `sentiments`. Finally, it adds this list as a new column named 'opinion_sentiment' to `test_df`.

Adjust the code according to your `get_sentiment` function and DataFrame structure as needed.
 Let me know if you need further clarification or assistance!

 **You**

/usr/local/lib/python3.10/dist-packages/ipykernel/ipkernel.py:283: DeprecationWarning:
 'should_run_async' will not call 'transform_cell' automatically in the future. Please pass the
 result to 'transformed_cell' argument and any exception that happens during the transform in
 'preprocessing_exc_tuple' in IPython 7.17 and above.
 and should_run_async(code)

```
-----  

AttributeError           Traceback (most recent call last)  

<ipython-input-20-52c6a18f4ebe> in <cell line: 3>()  

      3 for opinion in test_df['opinion']:  

      4     # Get sentiment for each opinion using the get_sentiment function  

----> 5     sentence_sentiment = get_sentiment(opinion)  

      6     sentiments.append(sentence_sentiment)  

      7  
  

<ipython-input-19-4834d17a7bc9> in get_sentiment(input_string)  

      15     ""  

      16     # calculate sentiment score with TextBlob, Afinn, NLTK+Vader  

--> 17     txt_blob = TextBlob(input_string.text).sentiment[0]  

      18     nltk_vader = sid.polarity_scores(input_string.text).get('compound')  

      19
```

AttributeError: 'str' object has no attribute 'text' ↓

Message ChatGPT



ChatGPT can make mistakes. Check important info.

 You

```
/usr/local/lib/python3.10/dist-packages/ipykernel/ipkernel.py:283: DeprecationWarning:
`should_run_async` will not call `transform_cell` automatically in the future. Please pass the
result to `transformed_cell` argument and any exception that happen during the transform in
`preprocessing_exc_tuple` in IPython 7.17 and above.
and should_run_async(code)
```

```
-----  
AttributeError           Traceback (most recent call last)  
<ipython-input-20-52c6a18f4ebe> in <cell line: 3>()  
      3 for opinion in test_df['opinion']: 3  
      4   # Get sentiment for each opinion using the get_sentiment function  
----> 5   sentence_sentiment = get_sentiment(opinion)  
      6   sentiments.append(sentence_sentiment)  
      7
```

```
<ipython-input-19-4834d17a7bc9> in get_sentiment(input_string)  
  15  "  
  16  #calculate sentiment score with TextBlob, Afinn, NLTK+Vader  
---> 17  txt_blob = TextBlob(input_string.text).sentiment[0]  
  18  nltk_vader = sid.polarity_scores(input_string.text).get('compound')  
  19
```

AttributeError: 'str' object has no attribute 'text'

 ChatGPT

The error message indicates that there is an `AttributeError` because you are trying to access the `'text'` attribute of a string object. It seems like `'input_string'` is a string instead of an object with a `'text'` attribute.

To resolve this, you need to ensure that the `'get_sentiment'` function receives the correct input. If `'input_string'` is a string representing the opinion text, you should directly pass it to the sentiment analysis functions without trying to access a `'text'` attribute.

Here's the corrected version of your code:

```
python                                         ⌂ Copy code  
  
# Assuming get_sentiment is a function that performs sentiment analysis on  
  
# Define the get_sentiment function  
def get_sentiment(input_string):  
    ...  
    This function calculates the sentiment score of an input string using ...  
    ...  
    # Calculate sentiment score with TextBlob, Afinn, NLTK+Vader  
    txt_blob = TextBlob(input_string).sentiment[0]  
    nltk_vader = sid.polarity_scores(input_string).get('compound')  
  
    # Return the sentiment scores  
    return txt_blob, nltk_vader  
  
# Loop through each opinion in test_df['opinion']  
sentiments = []  
for opinion in test_df['opinion']:  
    ↓
```

Message ChatGPT



ChatGPT can make mistakes. Check important info.

**You**

```
here's my get_sentiment: def get_sentiment(input_string):
    """
    Takes a string (sentence), and returns an averaged, three-model
    sentiment score, from -1 to 1. Where -1 is negative sentiment, 0 is neutral,
    and +1 is positive sentiment.
    """

    #calculate sentiment score with TextBlob, Afinn, NLTK+Vader
    txt_blob = TextBlob(input_string.text).sentiment[0]
    nltk_vader = sid.polarity_scores(input_string.text).get('compound')

    #TextBlob
    if txt_blob > 0.1:
        blob = 1
    elif txt_blob == 0 and txt_blob <= 0.1:
        blob = 0
    else: blob = -1
    #NLTK-Vader
    if nltk_vader > 0.1:
        vader = 1
    elif nltk_vader == 0 and nltk_vader <= 0.1:
        vader = 0
    else: vader = -1

    #Averaged return
    avg_score = (blob + vader) / 2
    return avg_score
```

**ChatGPT**

Thanks for sharing the `get_sentiment` function. Based on the provided function, you're calculating sentiment scores using TextBlob and NLTK+Vader, and then averaging the scores to get an overall sentiment score.

To integrate this function into your DataFrame and save the sentiment scores as a new column, you can use a similar approach as before. Here's how you can modify your code:

```
python
# Loop through each opinion in test_df['opinion']
sentiments = []
for opinion in test_df['opinion']:
    # Get sentiment for each opinion using the get_sentiment function
    sentence_sentiment = get_sentiment(opinion)
    sentiments.append(sentence_sentiment)

# Add the sentiment scores as a new column 'opinion_sentiment' to test_df
test_df['opinion_sentiment'] = sentiments

# Display the updated DataFrame
print(test_df)
```

This code will calculate the sentiment scores for each opinion in the 'opinion' column of `test_df` using the `get_sentiment` function and save the scores as a new column named 'opinion_sentiment'. Adjust the code according to your specific DataFrame structure and requirements.

Message ChatGPT



Ha

You
 /usr/local/lib/python3.10/dist-packages/ipykernel/ipkernel.py:283: DeprecationWarning:
 `should_run_async` will not call `transform_cell` automatically in the future. Please pass the
 result to `transformed_cell` argument and any exception that happen during the transform in
 `preprocessing_exc_tuple` in IPython 7.17 and above.
 and should_run_async(code)

```
-----
AttributeError           Traceback (most recent call last)
<ipython-input-20-52c6a18f4ebe> in <cell line: 3>()
      3 for opinion in test_df['opinion']:
      4     # Get sentiment for each opinion using the get_sentiment function
----> 5     sentence_sentiment = get_sentiment(opinion)
      6     sentiments.append(sentence_sentiment)
      7

<ipython-input-19-4834d17a7bc9> in get_sentiment(input_string)
      15     ""
      16     #calculate sentiment score with TextBlob, Afinn, NLTK+Vader
---> 17     txt_blob = TextBlob(input_string.text).sentiment[0]
      18     nltk_vader = sid.polarity_scores(input_string.text).get('compound')
      19
```

AttributeError: 'str' object has no attribute 'text'



ChatGPT

The error indicates that the `'input_string'` variable is a string object, and you're trying to access its `'text'` attribute, which doesn't exist. This suggests that `'input_string'` is already the text of the opinion, and there's no need to access any attribute.

To resolve this issue, you should directly pass the opinion text to your sentiment analysis functions without trying to access any attribute. Here's how you can modify your `'get_sentiment'` function:

```
python
Copy code

def get_sentiment(input_string):
    ...
    Takes a string (sentence), and returns an averaged, three-model
    sentiment score, from -1 to 1. Where -1 is negative sentiment, 0 is neutral
    and +1 is positive sentiment.
    ...
    #calculate sentiment score with TextBlob, Afinn, NLTK+Vader
    txt_blob = TextBlob(input_string).sentiment[0]
    nltk_vader = sid.polarity_scores(input_string).get('compound')

    #TextBlob
    if txt_blob > 0.1:
        blob = 1
    elif txt_blob == 0 and txt_blob <= 0.1:
        blob = 0
    else:
        blob = -1
    #NLTK-Vader
    if nltk_vader > 0.1:
```

Message ChatGPT

 Ha**You**

```
# Extract 'aspect', 'polarity', 'opinion', 'category' from the 'absa_result' column
test_df['aspect'] = absa_results.apply(lambda x: [item['aspect'] for item in x['Quadruples']])
test_df['polarity'] = absa_results.apply(lambda x: [item['polarity'] for item in x['Quadruples']])
test_df['opinion'] = absa_results.apply(lambda x: [item['opinion'] for item in x['Quadruples']])
test_df['category'] = absa_results.apply(lambda x: [item['category'] for item in x['Quadruples']])

# Print the DataFrame with the extracted columns
print(test_df[['aspect', 'polarity', 'opinion', 'category']])
```

**ChatGPT**

The code you provided is attempting to extract 'aspect', 'polarity', 'opinion', and 'category' from the 'absa_result' column in `test_df`. However, it seems there's a variable `absa_results` being used in the code without any definition. Presumably, `absa_results` should be the column containing the ABSA results.

If `absa_results` is indeed the column containing the ABSA results, you can directly apply the extraction operations without the need for the `apply` function. Here's how you can do it:

python

 Copy code

```
# Assuming 'absa_result' is the column containing ABSA results
test_df['aspect'] = test_df['absa_result'].apply(lambda x: [item['aspect']]
test_df['polarity'] = test_df['absa_result'].apply(lambda x: [item['polarity']]
test_df['opinion'] = test_df['absa_result'].apply(lambda x: [item['opinion']]
test_df['category'] = test_df['absa_result'].apply(lambda x: [item['category']]

# Print the DataFrame with the extracted columns
print(test_df[['aspect', 'polarity', 'opinion', 'category']])
```

Make sure that 'absa_result' column contains valid ABSA results in the expected format. If `absa_results` is a separate variable containing the ABSA results, then you need to adjust the code accordingly to use that variable.

Let me know if you need further assistance!

 Ha**You**

absa_results is actually a list

**ChatGPT**

If `absa_results` is a list containing ABSA results for each row in your DataFrame, you can iterate over the rows of the DataFrame and extract the desired information from each element of the `absa_results` list.

Here's how you can modify the code to handle `absa_results` as a list:

python

 Copy code

```
# Assuming 'absa_results' is a list containing ABSA results for each row in
```

Message ChatGPT

ChatGPT can make mistakes. Check important info.

 You

absa_results is actually a list

 ChatGPT

If `absa_results` is a list containing ABSA results for each row in your DataFrame, you can iterate over the rows of the DataFrame and extract the desired information from each element of the `absa_results` list.

Here's how you can modify the code to handle `absa_results` as a list:

python

 Copy code

```
# Assuming 'absa_results' is a list containing ABSA results for each row in the DataFrame
for i, absa_result in enumerate(absa_results):
    test_df.at[i, 'aspect'] = [item['aspect'] for item in absa_result['Quadruples']]
    test_df.at[i, 'polarity'] = [item['polarity'] for item in absa_result['Quadruples']]
    test_df.at[i, 'opinion'] = [item['opinion'] for item in absa_result['Quadruples']]
    test_df.at[i, 'category'] = [item['category'] for item in absa_result['Quadruples']]

# Print the DataFrame with the extracted columns
print(test_df[['aspect', 'polarity', 'opinion', 'category']])
```

In this code:

- We iterate over the rows of the DataFrame using `enumerate` to access both the index `i` and the corresponding ABSA result `absa_result`.
- For each row, we extract 'aspect', 'polarity', 'opinion', and 'category' from the `absa_result` and assign them to the corresponding columns in the DataFrame using `at` accessor.

Make sure that `absa_results` contains valid ABSA results for each row in the expected format.

Adjust the code according to your specific requirements and DataFrame structure.

Let me know if you need further assistance!

 You

Define aspects

```
defined_aspects = ['camera','battery','display','processor','storage','price','design']
```

```
from sklearn.metrics.pairwise import cosine_similarity
```

```
# Function to find the closest aspect for a given category
```

```
def find_closest_aspect(category, defined_aspects):
```

```
    category_embedding = sbert_model.encode([category])
```

```
    aspect_embeddings = sbert_model.encode(defined_aspects)
```

```
    similarities = cosine_similarity(category_embedding, aspect_embeddings)[0]
```

```
    closest_aspect_index = similarities.argmax()
```

```
    closest_aspect = defined_aspects[closest_aspect_index]
```

```
    return closest_aspect
```

```
# Apply the function to each category in test_df and save the result to a new column
```

```
test_df['defined_aspect'] = test_df['category'].apply(lambda x: find_closest_aspect(x, defined_aspects))
```



```
# Display the updated DataFrame
```

Message ChatGPT



ChatGPT can make mistakes. Check important info.

 You

```
# Define aspects
defined_aspects = ['camera','battery','display','processor','storage','price','design']

from sklearn.metrics.pairwise import cosine_similarity
# Function to find the closest aspect for a given category
def find_closest_aspect(category, defined_aspects):
    category_embedding = sbert_model.encode([category])
    aspect_embeddings = sbert_model.encode(defined_aspects)
    similarities = cosine_similarity(category_embedding, aspect_embeddings)[0]
    closest_aspect_index = similarities.argmax()
    closest_aspect = defined_aspects[closest_aspect_index]
    return closest_aspect

# Apply the function to each category in test_df and save the result to a new column
test_df['defined_aspect'] = test_df['category'].apply(lambda x: find_closest_aspect(x,
defined_aspects))

# Display the updated DataFrame
print(test_df) help me to adjust this code so that in the loop, after getting the similarities, if it's
>0.5, the code continues to extract the aspect, otherwise it returns "Others"
```

 ChatGPT

To adjust the code so that it checks if the highest similarity is greater than 0.5 before extracting the aspect, you can modify the `'find_closest_aspect'` function as follows:

```
python Copy code

# Function to find the closest aspect for a given category
def find_closest_aspect(category, defined_aspects, threshold=0.5):
    category_embedding = sbert_model.encode([category])
    aspect_embeddings = sbert_model.encode(defined_aspects)
    similarities = cosine_similarity(category_embedding, aspect_embeddings)

    # Check if the highest similarity is greater than the threshold
    if similarities.max() > threshold:
        closest_aspect_index = similarities.argmax()
        closest_aspect = defined_aspects[closest_aspect_index]
    else:
        closest_aspect = "Others"

    return closest_aspect

# Apply the function to each category in test_df and save the result to a new column
test_df['defined_aspect'] = test_df['category'].apply(lambda x: find_closest_aspect(x,
defined_aspects))

# Display the updated DataFrame
print(test_df)
```

In this modification:

- The `'find_closest_aspect'` function now takes an additional parameter `'threshold'`, which defaults to 0.5.

Message ChatGPT ↑

ChatGPT can make mistakes. Check important info.

You

```
# Iterate through each text to analyze sentiment
for text in texts:
    # TextBlob sentiment
    blob = TextBlob(text)
    tb_sentiment = blob.sentiment

    # VADER sentiment
    vader_sentiment = sid.polarity_scores(text)

    # BERT sentiment
    bert_result = nlp(text)

# Append results to data list
data.append({
    "Text": text,
    "TextBlob Sentiment": tb_sentiment[0],
    "VADER Compound Score": vader_sentiment['compound'],
    "BERT Sentiment": bert_result[0]['label'],
    "BERT Score": bert_result[0]['score'],
})
}) revise this code to use test_df['opinion'] as the input text, save the results as columns in the test_df
```



ChatGPT

To revise the provided code to use `test_df['opinion']` as the input text and save the sentiment analysis results as columns in `test_df`, you can iterate through the `test_df['opinion']` column and perform sentiment analysis for each opinion. Here's the revised code:

```
python Copy code

# Create empty lists to store sentiment analysis results
tb_sentiments = []
vader_compound_scores = []
bert_sentiments = []
bert_scores = []

# Iterate through each opinion in test_df['opinion'] to analyze sentiment
for opinion in test_df['opinion']:
    # TextBlob sentiment
    blob = TextBlob(opinion)
    tb_sentiment = blob.sentiment[0]
    tb_sentiments.append(tb_sentiment)

    # VADER sentiment
    vader_sentiment = sid.polarity_scores(opinion)['compound']
    vader_compound_scores.append(vader_sentiment)

    # BERT sentiment (assuming 'nlp' is your BERT sentiment analysis model)
    bert_result = nlp(opinion)
    bert_sentiment = bert_result[0]['label']
    bert_score = bert_result[0]['score']
    bert_scores.append(bert_score)
```

Message ChatGPT



ChatGPT can make mistakes. Check important info.

 You

give me the code to filter out all rows in which the column df['defined_aspect'] is Others

 ChatGPT

You can filter out all rows in which the column `df['defined_aspect']` contains "Others" using boolean indexing. Here's the code to achieve that:

```
python 
```

```
# Filter out rows where the column 'defined_aspect' is 'Others'
filtered_df = df[df['defined_aspect'] != 'Others']

# Display the filtered DataFrame
print(filtered_df)
```

This code selects only the rows where the value in the column 'defined_aspect' is not equal to "Others" and assigns the result to `filtered_df`.

Adjust the code according to your DataFrame variable name and structure. Let me know if you need further assistance!

 You

```
unique(df['defined_aspect'])
```

 ChatGPT

To get the unique values in the column `df['defined_aspect']`, you can use the `unique()` function provided by pandas. Here's how you can do it:

```
python 
```

```
unique_aspects = df['defined_aspect'].unique()
print(unique_aspects)
```

This code will print out the unique values present in the 'defined_aspect' column of your DataFrame `df`. Adjust the code according to your DataFrame variable name and structure. Let me know if you need further assistance!

 You

```
-----
KeyError          Traceback (most recent call last)
/usr/local/lib/python3.10/dist-packages/pandas/core/indexes/base.py in get_loc(self, key)
 3652     try:
-> 3653         return self._engine.get_loc(casted_key)
 3654     except KeyError as err:
```

4 frames

```
pandas/_libs/hashtable_class_helper.pxi in
pandas._libs.hashtable.PyObjectHashTable.getitem()
```



ChatGPT can make mistakes. Check important info.

 You

```
-----
KeyError           Traceback (most recent call last)
/usr/local/lib/python3.10/dist-packages/pandas/core/indexes/base.py in get_loc(self, key)
3652     try:
-> 3653         return self._engine.get_loc(casted_key)
3654     except KeyError as err:
```

4 frames
pandas/_libs/hashtable_class_helper.pxi in
pandas._libs.hashtable.PyObjectHashTable.get_item()

pandas/_libs/hashtable_class_helper.pxi in
pandas._libs.hashtable.PyObjectHashTable.get_item()

KeyError: 'defined_aspect'

The above exception was the direct cause of the following exception:

```
-----
KeyError           Traceback (most recent call last)
/usr/local/lib/python3.10/dist-packages/pandas/core/indexes/base.py in get_loc(self, key)
3653     return self._engine.get_loc(casted_key)
3654     except KeyError as err:
-> 3655         raise KeyError(key) from err
3656     except TypeError:
3657         # If we have a listlike key, _check_indexing_error will raise
```

KeyError: 'defined_aspect'



 ChatGPT

It seems that the column name 'defined_aspect' does not exist in your DataFrame. Please replace 'defined_aspect' with the correct column name that contains the aspects you want to extract.

For example, if the aspects are stored in a column named 'aspect', you would use:

```
python
unique_aspects = df['aspect'].unique()
print(unique_aspects)
```

Copy code

Replace 'aspect' with the actual column name in your DataFrame that contains the aspects. Let me know if you need further assistance!

 You

give me the code to filter out NULL value in test_df['opinion']

 ChatGPT

To filter out rows with NULL values in the 'opinion' column of the DataFrame `test_df`, you can use boolean indexing. Here's how you can do it:

```
python
# Filter out rows with NULL values in the 'opinion' column
```

Copy code

Message ChatGPT



ChatGPT can make mistakes. Check important info.

**ChatGPT**

To filter out rows with NULL values in the 'opinion' column of the DataFrame `test_df`, you can use boolean indexing. Here's how you can do it:

python

Copy code

```
# Filter out rows with NULL values in the 'opinion' column
filtered_df = test_df[test_df['opinion'].notnull()]

# Display the filtered DataFrame
print(filtered_df)
```

This code selects only the rows where the 'opinion' column is not null (``notnull()`` function) and assigns the result to `filtered_df`. Adjust the code according to your DataFrame variable name and structure. Let me know if you need further assistance!

**You**

why NULL value is still there

**ChatGPT**

If the code didn't filter out NULL values as expected, it's possible that the NULL values in your DataFrame are represented differently, such as 'None', 'nan', or an empty string. You can try additional methods to handle these cases:

1. **Handle 'None' Values:** If your NULL values are represented as 'None', you can filter them out using the `!= None` condition.
2. **Handle 'NaN' Values:** If your NULL values are represented as 'NaN' (Not a Number), you can use the `notna()` function.
3. **Handle Empty Strings:** If your NULL values are represented as empty strings, you can filter them out by checking if the string is not empty.

Here's how you can handle these cases:

python

Copy code

```
# Filter out rows with NULL, 'None', or 'nan' values in the 'opinion' column
filtered_df = test_df[test_df['opinion'].notnull() & (test_df['opinion'] != None) & (test_df['opinion'] != '')]

# Display the filtered DataFrame
print(filtered_df)
```

This code combines multiple conditions using the `&` operator to filter out rows with NULL, 'None', or 'nan' values, as well as empty strings, in the 'opinion' column. Adjust the code according to how NULL values are represented in your DataFrame. Let me know if you need further assistance!

**You**

no the NULL is still there



Message ChatGPT



ChatGPT can make mistakes. Check important info.

 **You**
no the NULL is still there

 **ChatGPT**
If the NULL values are still present after attempting to filter them out, it's possible that they are represented in a way that is not being caught by the filtering conditions. Here are a few more approaches you can try:

1. **Handle 'NaN' Values Explicitly:** In some cases, 'NaN' values may not be caught by the ``notnull()`` function. You can explicitly check for 'NaN' values using the ``pd.notna()`` function.
2. **Use `dropna()` :** Instead of filtering rows with NULL values, you can use the ``dropna()`` function to remove rows containing NULL values from the DataFrame altogether.

Here's how you can apply these approaches:

```
python 
```

```
# Method 1: Check for 'NaN' values explicitly
filtered_df = test_df[test_df['opinion'].notnull() & (test_df['opinion'] != 'NaN')]

# Method 2: Use dropna()
filtered_df = test_df.dropna(subset=['opinion'])

# Display the filtered DataFrame
print(filtered_df)
```

Try both methods to see if they effectively filter out the NULL values from the 'opinion' column. If the issue persists, there might be other factors at play, such as data type inconsistencies or unexpected values in the 'opinion' column. Let me know if you need further assistance!

 **You**
Explode lists into separate rows
`test_df = test_df.explode('aspect').explode('polarity').explode('opinion').explode('category')`

 **ChatGPT**
The code you provided explodes the lists in the columns 'aspect', 'polarity', 'opinion', and 'category' into separate rows, resulting in a new DataFrame with more rows. This is useful for situations where you have lists of values in a single cell that you want to separate into individual rows.

Here's a breakdown of what each ``explode()`` function does:

```
python 
```

```
test_df = test_df.explode('aspect') # Explode the 'aspect' column into separate rows
test_df = test_df.explode('polarity') # Explode the 'polarity' column into separate rows
test_df = test_df.explode('opinion') # Explode the 'opinion' column into separate rows
test_df = test_df.explode('category') # Explode the 'category' column into separate rows
```

After applying these ``explode()`` functions, each row in the DataFrame will contain only one value in

Message ChatGPT 

ChatGPT can make mistakes. Check important info.

You

😊\n', 'Quadruples': [{('aspect': 'iPhone 11', 'polarity': 'positive', 'opinion': 'fantastic', 'category': 'LAPTOP#QUALITY')}], {'text': 'Brought this dir my sons bday coming up he is going to be iver the moon\n', 'Quadruples': [{('aspect': 'dir', 'polarity': 'positive', 'opinion': 'NULL', 'category': 'SERVICE#GENERAL')}], {'text': 'I've purchased 2 recycled iPhone 14 years ago and a present for me last week, the iPhone 11 I bought is stunning no marks on it and battery power of 95% and it's recycle older phones at a price I can afford there is no way I can buy a brand new 11 I can't justify the price of a new 1, this way I get an iPhone 11 perfect in every way thankyou so much when I upgrade in afew years time I'd definitely buy another off you\n', 'Quadruples': [{('aspect': 'iPhone 11', 'polarity': 'positive', 'opinion': 'perfect', 'category': 'LAPTOP#GENERAL')}], {'text': 'Scratches on phone screen and battery not holding charge. Seller reusing to replace/refund, just said to charge and leave for 48 hours-funnily enough that didn't work either. AVIOD. This was a present for my daughters birthday and she was so upset. Now £200 out of pocket, also face recognition wouldn't work.\n', 'Quadruples': [{('aspect': 'phone screen', 'polarity': 'negative', 'opinion': 'Scratches', 'category': 'DISPLAY#QUALITY')}, {'aspect': 'battery', 'polarity': 'negative', 'opinion': 'NULL', 'category': 'BATTERY#QUALITY')}], {'text': 'This phone was delivered next day with Royal Mail. It works perfectly, no problems at all. Big savings on the cost of a new one. Great!!\n', 'Quadruples': [{('aspect': 'phone', 'polarity': 'positive', 'opinion': 'perfectly', 'category': 'LAPTOP#OPERATION_PERFORMANCE')}], {'text': 'The product was in good shape. Battery health was 99% and it had no scratches. I'm very impressed and it arrived before the estimated time. It's worth the price\n', 'Quadruples': [{('aspect': 'product', 'polarity': 'positive', 'opinion': 'good shape', 'category': 'LAPTOP#DESIGN_FEATURES')}, {'aspect': 'Battery health', 'polarity': 'positive', 'opinion': '99%', 'category': 'BATTERY#QUALITY')}, {'aspect': 'NULL', 'polarity': 'positive', 'opinion': 'worth the price', 'category': 'LAPTOP#PRICE')}], {'text': 'The iPhone 11 was a replacement phone for my father, I was a little hesitant to purchase online but took the gamble. I'm very impressed with the quality of the phone, not a mark on it, battery at 100% life, free screen saver thrown in. Couldn't fault it 🙌\n', 'Quadruples': [{('aspect': 'iPhone 11', 'polarity': 'positive', 'opinion': 'impressed', 'category': 'LAPTOP#GENERAL')}, {'aspect': 'battery', 'polarity': 'positive', 'opinion': '100%', 'category': 'BATTERY#OPERATION_PERFORMANCE')}], {'text': 'Phone arrived promptly but contactless and name drop not working so returned phone and received refund within a day. Faultless service\n', 'Quadruples': [{('aspect': 'Phone', 'polarity': 'positive', 'opinion': 'promptly', 'category': 'LAPTOP#GENERAL')}, {'aspect': 'name drop', 'polarity': 'negative', 'opinion': 'not working', 'category': 'SOFTWARE#OPERATION_PERFORMANCE')}], {'text': 'I was scared by the low reviews complaining about everything not working and so on but when i got the phone everything worked perfectly. great price and great phone!\n', 'Quadruples': [{('aspect': 'NULL', 'polarity': 'negative', 'opinion': 'crashed', 'category': 'LAPTOP#GENERAL')}, {'aspect': 'phone', 'polarity': 'positive', 'opinion': 'NULL', 'category': 'LAPTOP#PRICE')}], {'text': 'Better than I expected in all aspects\n', 'Quadruples': [{('aspect': 'NULL', 'polarity': 'positive', 'opinion': 'better', 'category': 'LAPTOP#GENERAL')}]}

```
/usr/local/lib/python3.10/dist-packages/ipykernel/ipkernel.py:283: DeprecationWarning: `should_run_async` will not call `transform_cell` automatically in the future. Please pass the result to `transformed_cell` argument and any exception that happen during the transform in `preprocessing_exc_tuple` in IPython 7.17 and above.
```

and should_run_async(code) I have a list like this, help me to extract each aspect with its corresponding opinion/polarity/category to a row



ChatGPT

To extract each aspect with its corresponding opinion, polarity, and category to a row, you can iterate over the list and create a new list of dictionaries where each dictionary represents a row with the required information. Here's how you can do it:

python



Copy code

Message ChatGPT



ChatGPT can make mistakes. Check important info.

 ChatGPT

To extract each aspect with its corresponding opinion, polarity, and category to a row, you can iterate over the list and create a new list of dictionaries where each dictionary represents a row with the required information. Here's how you can do it:

```
python
extracted_data = []

for item in data:
    text = item['text']
    quadruples = item['Quadruples']

    for quad in quadruples:
        aspect = quad['aspect']
        polarity = quad['polarity']
        opinion = quad['opinion']
        category = quad['category']

        extracted_data.append({
            'Text': text,
            'Aspect': aspect,
            'Polarity': polarity,
            'Opinion': opinion,
            'Category': category
        })

# Convert the list of dictionaries to a DataFrame
extracted_df = pd.DataFrame(extracted_data)

# Display the DataFrame
print(extracted_df)
```

This code iterates over each item in the list and extracts the required information from the 'Quadruples' list. It then appends this information to a new list of dictionaries. Finally, it converts the list of dictionaries to a DataFrame.

Let me know if you need further assistance!

 You

```
review_title \
0      Really happy
1  IMO, the best premium mobile phone by far...
2      Quick,Great camera
3      Nice design
4  The screen freezes and I charge the battery ev...
5      Fabulous!
6      Just does everything
7      Great phone
8  It's Apple so you expect the best
9      Good phone
10     Lovely phone
```



Message ChatGPT



ChatGPT can make mistakes. Check important info.

**ChatGPT**

To transform the data into single-value rows

**You**

How do I transform the dataframe I gave you into single-value rows?

**ChatGPT**

To transform the dataframe into single-value rows, you can iterate through each row and combine the values of the lists into single values. Here's how you can do it:

```
python Copy code
import pandas as pd

# Your dataframe
df = pd.DataFrame({ 'review_title': ['Really happy', 'IMO, the best premium'],
                     'review_text': ['Upgraded from an iPhone 13 Pro Max. No',
                     'review_rating': [5, 5, 5],
                     'review_date': ['09-Apr-24', '20-Feb-24', '21-Apr-24'],
                     'overall_rating': [4.5, 4.5, 4.5],
                     'aspect': [['iPhone 13 Pro Max'], ['device', 'camera /',
                     'polarity': ['positive', 'positive', 'positive'],
                     'opinion': ['great', 'superb', 'superb', 'excellent'],
                     'category': [['LAPTOP#OPERATION_PERFORMANCE'], ['LAPTOP#OPERATION_PERFORMANCE']]]

# Create an empty list to store the transformed rows
transformed_rows = []

# Iterate through each row
for index, row in df.iterrows():
    # Iterate through each aspect, polarity, opinion, and category
    for i in range(len(row['aspect'])):
        # Create a new row with combined values
        new_row = {
            'review_title': row['review_title'],
            'review_text': row['review_text'],
            'review_rating': row['review_rating'],
            'review_date': row['review_date'],
            'overall_rating': row['overall_rating'],
            'aspect': row['aspect'][i],
            'polarity': row['polarity'][i],
            'opinion': row['opinion'][i],
            'category': row['category'][i]
        }
        # Append the new row to the list
        transformed_rows.append(new_row)
    # Create a new dataframe from the transformed rows list
    ↓
```

Message ChatGPT



ChatGPT can make mistakes. Check important info.

 You

no I mean transform the dataframe I gave you into single-value-row dataframe?

 ChatGPT

Got it. To transform the dataframe into a single-value-row dataframe, you can concatenate the lists within each column into single strings. Here's how you can do it:

```
python  Copy code

import pandas as pd

# Your dataframe
df = pd.DataFrame({ 'review_title': ['Really happy', 'IMO, the best premium'],
                     'review_text': ['Upgraded from an iPhone 13 Pro Max. No',
                     'review_rating': [5, 5, 5],
                     'review_date': ['09-Apr-24', '20-Feb-24', '21-Apr-24'],
                     'overall_rating': [4.5, 4.5, 4.5],
                     'aspect': [['iPhone 13 Pro Max'], ['device', 'camera /',
                     'polarity': ['positive', ['positive', 'positive', 'posi',
                     'opinion': ['great', ['superb', 'superb', 'excellent', 'exce
                     'category': [['LAPTOP#OPERATION_PERFORMANCE'], ['LAPTOP#OPERA
                     ')]

# Concatenate lists into single strings
df['aspect'] = df['aspect'].apply(lambda x: ', '.join(x) if isinstance(x, list) else x)
df['polarity'] = df['polarity'].apply(lambda x: ', '.join(x) if isinstance(x, list) else x)
df['opinion'] = df['opinion'].apply(lambda x: ', '.join(x) if isinstance(x, list) else x)
df['category'] = df['category'].apply(lambda x: ', '.join(x) if isinstance(x, list) else x)

# Create a single-row dataframe
single_row_df = df.head(1)

# Display the single-row dataframe
print(single_row_df)
```

This code will concatenate the lists within each column into single strings. Then, it will select the first row of the dataframe to create a single-row dataframe.

 You

I dont want to concatenate the list into a single string, I want to split each string in the list into a row for each column, but dont do all combination, the combination of columns will follow the index of the string in the list. For example, in row 8, aspect is currently [camera, Battery life], the polarity is [positive, positive], opinion is [better, better]. So I want to split this row 8 into 2 rows, 1 row with aspect being camera, polarity being positive and opinion being better while the other row with aspect being battery life, polarity being positive and opinion being better, with all the other columns keeping the same values as they originally arwe

 ChatGPT

 I see, you want to explode each list in the dataframe column into separate rows while keeping the other

Message ChatGPT 

ChatGPT can make mistakes. Check important info.

 You

I dont want to concatenate the list into a single string, I want to split each string in the list into a row for each column, but dont do all combination, the combination of columns will follow the index of the string in the list. For example, in row 8, aspect is currently [camera, Battery life], the polarity is [positive, positive], opinion is [better, better]. So I want to split this row 8 into 2 rows, 1 row with aspect being camera, polarity being positive and opinion being better while the other row with aspect being battery life, polarity being positive and opinion being better, with all the other columns keeping the same values as they originally arwe

 ChatGPT

I see, you want to explode each list in the dataframe column into separate rows while keeping the other columns intact. Here's how you can do it:

```
python
import pandas as pd

# Your dataframe
df = pd.DataFrame({ 'review_title': ['Really happy', 'IMO, the best premium'],
                     'review_text': ['Upgraded from an iPhone 13 Pro Max. No',
                                    'longer'],
                     'review_rating': [5, 5, 5],
                     'review_date': ['09-Apr-24', '20-Feb-24', '21-Apr-24'],
                     'overall_rating': [4.5, 4.5, 4.5],
                     'aspect': [['iPhone 13 Pro Max'], ['device', 'camera /'],
                                ['polarity': ['positive', ['positive', 'positive', 'posi'],
                                'opinion': ['great', ['superb', 'superb', 'excellent'],
                                'category': [['LAPTOP#OPERATION_PERFORMANCE'], ['LAPTOP#OPERATION_PERFORMANCE'],
                                              ['LAPTOP#OPERATION_PERFORMANCE']]])]

# Define a function to explode lists into separate rows
def explode_lists(df, col):
    exploded_df = df.drop(col, axis=1).join(df[col].explode()).reset_index()
    return exploded_df

# Apply the function to explode each column with lists
columns_to_explode = ['aspect', 'polarity', 'opinion', 'category']
for col in columns_to_explode:
    df = explode_lists(df, col)

# Display the exploded dataframe
print(df)
```

This code will explode each list in the specified columns into separate rows while keeping the other columns intact.

 You

test_df.index(1)

 ChatGPT

The `index(1)` method in Python's Pandas library is used to get the index (row label) of the first

Message ChatGPT

ChatGPT can make mistakes. Check important info.

 You
test_df.index(1)

 ChatGPT
The `index()` method in Python's Pandas library is used to get the index (row label) of the first occurrence of the specified value. However, it seems you're trying to use it incorrectly.

To get the index of a specific row based on its position, you can directly access the index using square brackets notation `[]`. Here's how you can get the index of the row at position 1 (the second row):

```
python 
index_at_position_1 = test_df.index[1]
print(index_at_position_1)
```

Replace `test_df` with the name of your DataFrame. This will give you the index label of the row at position 1.

 You
I want to print all rows with index being 1

 ChatGPT
You can use boolean indexing to filter rows based on their index values. Here's how you can print all rows where the index is 1:

```
python 
rows_with_index_1 = test_df[test_df.index == 1]
print(rows_with_index_1)
```

This will print all rows where the index is 1. Replace `test_df` with the name of your DataFrame.

 You
I see, you want to explode each list in the dataframe column into separate rows while keeping the other columns intact. Here's how you can do it:

explode here is not correct. I told you I dont want all possible combination of each string in each list. I just want the combination following order, like the first string in the aspect column list should be combined with the first string in the polarity/opinion/category columns

 ChatGPT
Got it. You can achieve this by using the `apply` function along with `pd.Series` to split each list into separate rows while keeping the other columns unchanged. Here's how you can do it:

```
python 
import pandas as pd
# Your DataFrame
```

Message ChatGPT 

ChatGPT can make mistakes. Check important info.

 **You**

I see, you want to explode each list in the dataframe column into separate rows while keeping the other columns intact. Here's how you can do it:

explode here is not correct. I told you I dont want all possible combination of each string in each list. I just want the combination following order, like the first string in the aspect column list should be combined with the first string in the polarity/opinion/category columns

 **ChatGPT**

Got it. You can achieve this by using the `apply` function along with `pd.Series` to split each list into separate rows while keeping the other columns unchanged. Here's how you can do it:

```
python Copy code

import pandas as pd

# Your DataFrame
# test_df = ...

# Define a function to explode each list in a row into separate rows
def explode_lists(row):
    # Get the length of the lists in each column
    max_len = max(map(len, row))
    # Create a DataFrame with the exploded lists
    exploded_df = pd.DataFrame()
    for col in row.index:
        # If the column contains a list, explode it
        if isinstance(row[col], list):
            # Pad the list with NaNs to match the length of the longest list
            padded_list = row[col] + [None] * (max_len - len(row[col]))
            # Add the exploded list as a new column
            exploded_df[col] = padded_list
        else:
            # If the column doesn't contain a list, repeat its value for each row
            exploded_df[col] = [row[col]] * max_len
    return exploded_df

# Apply the function to each row of the DataFrame
exploded_rows = test_df.apply(explode_lists, axis=1)

# Concatenate the resulting DataFrames into a single DataFrame
exploded_df = pd.concat(exploded_rows.tolist(), ignore_index=True)

# Print the resulting DataFrame
print(exploded_df)
```

This code will split each list in the DataFrame into separate rows, following the order of elements in each list.

 **You**

Message ChatGPT ↑

ChatGPT can make mistakes. Check important info.



You

```

TypeError           Traceback (most recent call last)
<ipython-input-65-4640cf395f8f> in <cell line: 23>()
     21
     22 # Apply the function to each row of the DataFrame
--> 23 exploded_rows = test_df.apply(explode_lists, axis=1)
     24
     25 # Concatenate the resulting DataFrames into a single DataFrame

4 frames
<ipython-input-65-4640cf395f8f> in explode_lists(row)
      5 def explode_lists(row):
      6     # Get the length of the lists in each column
----> 7     max_len = max(map(len, row))
      8     # Create a DataFrame with the exploded lists
      9     exploded_df = pd.DataFrame()

TypeError: object of type 'int' has no len()

```



ChatGPT

My apologies for the oversight. It seems there was an issue with the code. Let's adjust it to properly handle the lists within each row. We'll iterate over the columns instead of the rows. Here's the corrected version:

```

python
Copy code

import pandas as pd

# Your DataFrame
# test_df = ...

# Define a function to explode each list in a row into separate rows
def explode_lists(row):
    # Create a dictionary to hold the exploded values
    exploded_dict = {}
    for col in test_df.columns:
        if isinstance(row[col], list):
            # Explode the list and add it to the dictionary
            for i, val in enumerate(row[col]):
                exploded_dict[f"{col}_{i}"] = val
        else:
            # If the value is not a list, keep it as is
            exploded_dict[col] = row[col]
    return exploded_dict

# Apply the function to each row of the DataFrame
exploded_rows = test_df.apply(explode_lists, axis=1)

# Create a DataFrame from the exploded rows
exploded_df = pd.DataFrame(exploded_rows.tolist())
↓
# Print the resulting DataFrame

```

Message ChatGPT



ChatGPT can make mistakes. Check important info.

 You

I dont want to split the list into different columns, I want different rows

 ChatGPT

Apologies for misunderstanding. If you want to split each row into multiple rows based on the elements in the lists while keeping the other columns intact, you can achieve it like this:

```
python Copy code

import pandas as pd

# Your DataFrame
# test_df = ...

# Initialize an empty list to store the new rows
new_rows = []

# Iterate over each row in the DataFrame
for _, row in test_df.iterrows():
    # Get the maximum length among the lists in the row
    max_len = max(map(len, row[['aspect', 'polarity', 'opinion', 'category']]))

    # Create a new row for each index up to the max length
    for i in range(max_len):
        new_row = {
            'review_title': row['review_title'],
            'review_text': row['review_text'],
            'review_rating': row['review_rating'],
            'review_date': row['review_date'],
            'overall_rating': row['overall_rating']
        }

        # Get the value at each index for aspect, polarity, opinion, and category
        new_row['aspect'] = row['aspect'][i] if i < len(row['aspect']) else None
        new_row['polarity'] = row['polarity'][i] if i < len(row['polarity']) else None
        new_row['opinion'] = row['opinion'][i] if i < len(row['opinion']) else None
        new_row['category'] = row['category'][i] if i < len(row['category']) else None

        # Append the new row to the list
        new_rows.append(new_row)

# Create a DataFrame from the new rows
new_df = pd.DataFrame(new_rows)

# Print the resulting DataFrame
print(new_df)
```

This code will create a new DataFrame where each row is split into multiple rows based on the elements in the lists, while keeping the other columns intact.

Message ChatGPT ↑

ChatGPT can make mistakes. Check important info.

 You

add a filter to make sure this loop only works on strings within a list, not slicing the string into letter,

 ChatGPT

Certainly! You can add a check to ensure that only lists are processed, like this:

```
python 
```

```

import pandas as pd

# Your DataFrame
# test_df = ...

# Initialize an empty list to store the new rows
new_rows = []

# Iterate over each row in the DataFrame
for _, row in test_df.iterrows():
    # Filter out non-list values
    aspect_list = row['aspect'] if isinstance(row['aspect'], list) else [row['aspect']]
    polarity_list = row['polarity'] if isinstance(row['polarity'], list) else [row['polarity']]
    opinion_list = row['opinion'] if isinstance(row['opinion'], list) else [row['opinion']]
    category_list = row['category'] if isinstance(row['category'], list) else [row['category']]

    # Get the maximum length among the lists in the row
    max_len = max(map(len, [aspect_list, polarity_list, opinion_list, category_list]))

    # Create a new row for each index up to the max length
    for i in range(max_len):
        new_row = {
            'review_title': row['review_title'],
            'review_text': row['review_text'],
            'review_rating': row['review_rating'],
            'review_date': row['review_date'],
            'overall_rating': row['overall_rating']
        }

        # Get the value at each index for aspect, polarity, opinion, and category
        new_row['aspect'] = aspect_list[i] if i < len(aspect_list) else None
        new_row['polarity'] = polarity_list[i] if i < len(polarity_list) else None
        new_row['opinion'] = opinion_list[i] if i < len(opinion_list) else None
        new_row['category'] = category_list[i] if i < len(category_list) else None

        # Append the new row to the list
        new_rows.append(new_row)

# Create a DataFrame from the new rows
new_df = pd.DataFrame(new_rows) 

```

Message ChatGPT 

ChatGPT can make mistakes. Check important info.

 You

what if I want to add these new rows into the test_df

 ChatGPT

You can concatenate the new DataFrame with the original `test_df` using the `pd.concat()` function. Here's how you can do it:

```
python Copy code
import pandas as pd

# Initialize an empty list to store the new rows
new_rows = []

# Iterate over each row in the DataFrame
for _, row in test_df.iterrows():
    # Filter out non-list values
    aspect_list = row['aspect'] if isinstance(row['aspect'], list) else [row['aspect']]
    polarity_list = row['polarity'] if isinstance(row['polarity'], list) else [row['polarity']]
    opinion_list = row['opinion'] if isinstance(row['opinion'], list) else [row['opinion']]
    category_list = row['category'] if isinstance(row['category'], list) else [row['category']]

    # Get the maximum length among the lists in the row
    max_len = max(map(len, [aspect_list, polarity_list, opinion_list, category_list]))

    # Create a new row for each index up to the max length
    for i in range(max_len):
        new_row = {
            'review_title': row['review_title'],
            'review_text': row['review_text'],
            'review_rating': row['review_rating'],
            'review_date': row['review_date'],
            'overall_rating': row['overall_rating']
        }

        # Get the value at each index for aspect, polarity, opinion, and category
        new_row['aspect'] = aspect_list[i] if i < len(aspect_list) else None
        new_row['polarity'] = polarity_list[i] if i < len(polarity_list) else None
        new_row['opinion'] = opinion_list[i] if i < len(opinion_list) else None
        new_row['category'] = category_list[i] if i < len(category_list) else None

        # Append the new row to the list
        new_rows.append(new_row)

# Create a DataFrame from the new rows
new_df = pd.DataFrame(new_rows)

# Concatenate the new DataFrame with the original test_df
test_df = pd.concat([test_df, new_df], ignore_index=True)
```

Message ChatGPT


ChatGPT can make mistakes. Check important info.

 You

↪ Certainly! You can add a check to ensure that only lists are processed, like this:

in this code, in the loop also include other columns with the values unchanged to maintain the whole df



ChatGPT

```
python Copy code

import pandas as pd

# Initialize an empty list to store the new rows
new_rows = []

# Iterate over each row in the DataFrame
for _, row in test_df.iterrows():
    # Filter out non-list values
    aspect_list = row['aspect'] if isinstance(row['aspect'], list) else [row['aspect']]
    polarity_list = row['polarity'] if isinstance(row['polarity'], list) else [row['polarity']]
    opinion_list = row['opinion'] if isinstance(row['opinion'], list) else [row['opinion']]
    category_list = row['category'] if isinstance(row['category'], list) else [row['category']]

    # Get the maximum length among the lists in the row
    max_len = max(len(aspect_list), len(polarity_list), len(opinion_list), len(category_list))

    # Create a new row for each index up to the max length
    for i in range(max_len):
        new_row = row.copy() # Copy the original row

        # Get the value at each index for aspect, polarity, opinion, and category
        new_row['aspect'] = aspect_list[i] if i < len(aspect_list) else None
        new_row['polarity'] = polarity_list[i] if i < len(polarity_list) else None
        new_row['opinion'] = opinion_list[i] if i < len(opinion_list) else None
        new_row['category'] = category_list[i] if i < len(category_list) else None

        # Append the new row to the list
        new_rows.append(new_row)

# Create a DataFrame from the new rows
new_df = pd.DataFrame(new_rows)

# Concatenate the new DataFrame with the original test_df
test_df = pd.concat([test_df, new_df], ignore_index=True)

# Print the updated DataFrame
print(test_df)
```

This will preserve all other columns in the DataFrame while adding the new rows based on the lists in the `aspect`, `polarity`, `opinion`, and `category` columns.

Message ChatGPT ↑

ChatGPT can make mistakes. Check important info.



You

no but doing like this you lost which values in other columns are associated with with row in the new_df??



ChatGPT

You're right; to preserve the association between the new rows and the original rows in other columns, you can add identifiers to each row indicating the original index. Here's how you can modify the code to achieve that:

python

Copy code

```

import pandas as pd

# Initialize an empty list to store the new rows
new_rows = []

# Iterate over each row in the DataFrame
for index, row in test_df.iterrows():
    # Filter out non-list values
    aspect_list = row['aspect'] if isinstance(row['aspect'], list) else [row['aspect']]
    polarity_list = row['polarity'] if isinstance(row['polarity'], list) else [row['polarity']]
    opinion_list = row['opinion'] if isinstance(row['opinion'], list) else [row['opinion']]
    category_list = row['category'] if isinstance(row['category'], list) else [row['category']]

    # Get the maximum length among the lists in the row
    max_len = max(map(len, [aspect_list, polarity_list, opinion_list, category_list]))

    # Create a new row for each index up to the max length
    for i in range(max_len):
        new_row = row.copy() # Copy the original row

        # Get the value at each index for aspect, polarity, opinion, and category
        new_row['aspect'] = aspect_list[i] if i < len(aspect_list) else None
        new_row['polarity'] = polarity_list[i] if i < len(polarity_list) else None
        new_row['opinion'] = opinion_list[i] if i < len(opinion_list) else None
        new_row['category'] = category_list[i] if i < len(category_list) else None

        # Add identifier columns to preserve association
        new_row['original_index'] = index

        # Append the new row to the list
        new_rows.append(new_row)

# Create a DataFrame from the new rows
new_df = pd.DataFrame(new_rows)

# Concatenate the new DataFrame with the original test_df
test_df = pd.concat([test_df, new_df], ignore_index=True)

```

Message ChatGPT



ChatGPT can make mistakes. Check important info.

You

Yes, but you didn't use that column at all

ChatGPT

You're correct, my mistake. Let's utilize the `original_index` column to preserve the association between the new rows and the original rows. Here's the modified code to include that:

```
python  Copy code

import pandas as pd

# Initialize an empty list to store the new rows
new_rows = []

# Iterate over each row in the DataFrame
for index, row in test_df.iterrows():
    # Filter out non-list values
    aspect_list = row['aspect'] if isinstance(row['aspect'], list) else [row['aspect']]
    polarity_list = row['polarity'] if isinstance(row['polarity'], list) else [row['polarity']]
    opinion_list = row['opinion'] if isinstance(row['opinion'], list) else [row['opinion']]
    category_list = row['category'] if isinstance(row['category'], list) else [row['category']]

    # Get the maximum length among the lists in the row
    max_len = max(len(aspect_list), len(polarity_list), len(opinion_list), len(category_list))

    # Create a new row for each index up to the max length
    for i in range(max_len):
        new_row = row.copy() # Copy the original row

        # Get the value at each index for aspect, polarity, opinion, and category
        new_row['aspect'] = aspect_list[i] if i < len(aspect_list) else None
        new_row['polarity'] = polarity_list[i] if i < len(polarity_list) else None
        new_row['opinion'] = opinion_list[i] if i < len(opinion_list) else None
        new_row['category'] = category_list[i] if i < len(category_list) else None

        # Add identifier columns to preserve association
        new_row['original_index'] = index

        # Append the new row to the list
        new_rows.append(new_row)

# Create a DataFrame from the new rows
new_df = pd.DataFrame(new_rows)

# Concatenate the new DataFrame with the original test_df
test_df = pd.concat([test_df, new_df], ignore_index=True)

# Print the updated DataFrame
print(test_df)
```

Message ChatGPT

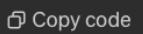
ChatGPT can make mistakes. Check important info.

 You

no, I mean when concatenate new_df with test_df, if u dont specify how can it know which row should be concatenated with which????

 ChatGPT

You're right; I apologize for misunderstanding. To maintain the association between the new rows and the original ones, we need to include an identifier column in both DataFrames. We can achieve this by adding an `'original_index'` column to the new DataFrame, which will store the index of the original row from which the new row was derived. Then, when concatenating `'new_df'` with `'test_df'`, we can specify `'original_index'` as the column to use for alignment. Here's how we can do it:

```
python 
```

```

import pandas as pd

# Initialize an empty list to store the new rows
new_rows = []

# Iterate over each row in the DataFrame
for index, row in test_df.iterrows():
    # Filter out non-list values
    aspect_list = row['aspect'] if isinstance(row['aspect'], list) else [row['aspect']]
    polarity_list = row['polarity'] if isinstance(row['polarity'], list) else [row['polarity']]
    opinion_list = row['opinion'] if isinstance(row['opinion'], list) else [row['opinion']]
    category_list = row['category'] if isinstance(row['category'], list) else [row['category']]

    # Get the maximum length among the lists in the row
    max_len = max(map(len, [aspect_list, polarity_list, opinion_list, category_list]))

    # Create a new row for each index up to the max length
    for i in range(max_len):
        new_row = row.copy() # Copy the original row

        # Get the value at each index for aspect, polarity, opinion, and category
        new_row['aspect'] = aspect_list[i] if i < len(aspect_list) else None
        new_row['polarity'] = polarity_list[i] if i < len(polarity_list) else None
        new_row['opinion'] = opinion_list[i] if i < len(opinion_list) else None
        new_row['category'] = category_list[i] if i < len(category_list) else None

        # Add identifier column to preserve association
        new_row['original_index'] = index

        # Append the new row to the list
        new_rows.append(new_row)

# Create a DataFrame from the new rows
new_df = pd.DataFrame(new_rows)

# Concatenate the new DataFrame with the original test_df, using original_index as the key
test_df = pd.concat([test_df, new_df], ignore_index=True, sort=False)

```

Message ChatGPT 

ChatGPT can make mistakes. Check important info.

You

```
# Remove "LAPTOP" or "#" from each entry in the 'category' column using regex
test_df['category'] = test_df['category'].str.replace("LAPTOP#", "", regex=True) # Remove
# at the beginning
test_df['category'] = test_df['category'].str.replace("#", "", regex=True) # Remove any other #
```

ChatGPT

The code you provided is correct for removing "LAPTOP#" and "#" from each entry in the 'category' column using regular expressions. It first removes "LAPTOP#" at the beginning of each entry and then removes any remaining "#" symbols. This will effectively clean up the category values in the DataFrame. If you have any questions or need further assistance, feel free to ask!

You

how to extract sentiment score using BERT

ChatGPT

To extract sentiment scores using BERT, you typically use a pre-trained BERT model fine-tuned on a sentiment analysis task. Here's a general outline of the steps:

1. **Load Pre-trained BERT Model:** You need to load a pre-trained BERT model. You can use libraries like Hugging Face's `'transformers'` library to load pre-trained BERT models.
2. **Tokenization:** Tokenize your text data using the BERT tokenizer. This step involves converting your text input into tokens understandable by the BERT model.
3. **Padding and Truncation:** Ensure that all tokenized sequences have the same length by padding or truncating them.
4. **Model Inference:** Pass the tokenized input through the BERT model to get the hidden representations.
5. **Classification Head:** Add a classification head (typically a fully connected layer) on top of the BERT model to predict sentiment scores.
6. **Inference and Post-processing:** Perform inference to get the sentiment scores for each input.

Here's an example code snippet using Hugging Face's `'transformers'` library to extract sentiment scores using a pre-trained BERT model:

```
python  Copy code

from transformers import BertTokenizer, BertForSequenceClassification
import torch

# Load pre-trained BERT model and tokenizer
model_name = 'bert-base-uncased'
tokenizer = BertTokenizer.from_pretrained(model_name)
model = BertForSequenceClassification.from_pretrained(model_name)

# Sample text
text = "This is a positive review." 
```

Tokenize and prepare input inputs = tokenizer(text, return_tensors='pt', padding=True, truncation=True)

Message ChatGPT

ChatGPT can make mistakes. Check important info.

**You**

```
# Iterating over the rows of the DataFrame and extract the desired information from each
element of absa_results
for i, absa_result in enumerate(absa_results):
    df.at[i, 'aspect'] = [item['aspect'] for item in absa_result['Quadruples']]
    df.at[i, 'polarity'] = [item['polarity'] for item in absa_result['Quadruples']]
    df.at[i, 'opinion'] = [item['opinion'] for item in absa_result['Quadruples']]
    df.at[i, 'category'] = [item['category'] for item in absa_result['Quadruples']] that error is from
this
```

**ChatGPT**

The error message suggests that the assignment operation is expecting equal lengths of keys and values. However, the length of the keys and values might not be the same, causing this ValueError.

To resolve this, you can try appending the extracted information from each element of `absa_results` to separate lists and then assign these lists to the DataFrame columns. This way, you ensure that the lengths of the keys and values are consistent.

Here's how you can modify the code:

```
python Copy code

# Initialize empty lists to store extracted information
aspects = []
polarities = []
opinions = []
categories = []

# Iterating over the rows of the DataFrame and extract the desired information
for absa_result in absa_results:
    for item in absa_result['Quadruples']:
        aspects.append(item['aspect'])
        polarities.append(item['polarity'])
        opinions.append(item['opinion'])
        categories.append(item['category'])

# Assign lists to DataFrame columns
df['aspect'] = aspects
df['polarity'] = polarities
df['opinion'] = opinions
df['category'] = categories
```

This will ensure that each list has the same length, avoiding the ValueError. Adjust the code as needed based on your DataFrame and data structure.

**You**

[{"text": "upgraded from an iphone 13 pro max normally get a contract but thought i d just buy the phone off amazon and then get a 12 month contract works great and really good camera no faults at all taken a while to be delivered as i think the first one got stolen in an amazon depot", "down": "\u2193"}, {"text": "'Quadruples': [{"aspect": "phone",

Message ChatGPT \u2191

ChatGPT can make mistakes. Check important info.

 You
add it to my original df dataset

 ChatGPT
To add the extracted information to your original DataFrame (`df`), you can use the `pd.concat()` function to concatenate the original DataFrame with the DataFrame created from the extracted information. Here's how you can do it:

```
python Copy code
import pandas as pd

# Initialize empty lists to store extracted information
aspects = []
polarities = []
opinions = []
categories = []

# Iterate over the list of dictionaries in absa_results
for absa_result in absa_results:
    # Iterate over the 'Quadruples' key in each dictionary
    for quadruple in absa_result['Quadruples']:
        # Extract aspect, polarity, opinion, and category from each quadruple
        aspects.append(quadruple['aspect'])
        polarities.append(quadruple['polarity'])
        opinions.append(quadruple['opinion'])
        categories.append(quadruple['category'])

# Create a DataFrame from the extracted information
extracted_df = pd.DataFrame({
    'aspect': aspects,
    'polarity': polarities,
    'opinion': opinions,
    'category': categories
})

# Concatenate the original DataFrame with the DataFrame created from the extracted information
df = pd.concat([df, extracted_df], ignore_index=True)

# Display the updated DataFrame
print(df)
```

This will add the extracted information to your original DataFrame `df`, creating additional rows.
Adjust the code as needed based on your DataFrame and data structure.



 You
don't use concat, why not use this logic: # Iterating over the rows of the DataFrame and extract the desired information from each element of absa_results
for i, absa_result in enumerate(absa_results): 
df.at[i, 'aspect'] = [item['aspect'] for item in absa_result['Quadruples']]

Message ChatGPT ↑

ChatGPT can make mistakes. Check important info.

 You

```
don't use concat, why not use this logic: # Iterating over the rows of the DataFrame and extract the desired information from each element of absa_results
for i, absa_result in enumerate(absa_results):
    df.at[i, 'aspect'] = [item['aspect'] for item in absa_result['Quadruples']]
    df.at[i, 'polarity'] = [item['polarity'] for item in absa_result['Quadruples']]
    df.at[i, 'opinion'] = [item['opinion'] for item in absa_result['Quadruples']]
    df.at[i, 'category'] = [item['category'] for item in absa_result['Quadruples']]
```

 ChatGPT

You're right, you can directly iterate over the rows of the DataFrame and extract the desired information from each element of `absa_results`. Here's how you can do it using your logic:

python

 Copy code

```
# Iterating over the rows of the DataFrame and extract the desired information
for i, absa_result in enumerate(absa_results):
    # Initialize lists to store extracted information for the current row
    aspects = []
    polarities = []
    opinions = []
    categories = []

    # Iterate over the 'Quadruples' key in the current absa_result
    for quadruple in absa_result['Quadruples']:
        # Extract aspect, polarity, opinion, and category from each quadruple
        aspects.append(quadruple['aspect'])
        polarities.append(quadruple['polarity'])
        opinions.append(quadruple['opinion'])
        categories.append(quadruple['category'])

    # Assign the extracted information to the corresponding rows in the DataFrame
    df.at[i, 'aspect'] = aspects
    df.at[i, 'polarity'] = polarities
    df.at[i, 'opinion'] = opinions
    df.at[i, 'category'] = categories
```

This code will iterate over each row of the DataFrame and extract the desired information from the corresponding element of `absa_results`, then assign it to the appropriate columns in the DataFrame.

 You

still error

 ChatGPT

I see, it seems there might be an issue with the structure of your DataFrame or the format of `absa_results`. Let's debug it step by step:

1. **Check DataFrame Structure:** Make sure your DataFrame `df` has columns named 'aspect',

Message ChatGPT





ChatGPT

Thanks for providing the `'absa_results'`. Now, let's iterate over each element of `'absa_results'` and extract the information into the DataFrame `'df'`:

```
python Copy code
# Iterate over the rows of the DataFrame and extract the desired information
for i, absa_result in enumerate(absa_results):
    for quad in absa_result['Quadruples']:
        df = df.append({'aspect': quad['aspect'],
                        'polarity': quad['polarity'],
                        'opinion': quad['opinion'],
                        'category': quad['category']},
                       ignore_index=True)
```

This loop iterates over each element of ``absa_results``, then iterates over the list of dictionaries under the key `'`Quadruples`'` for each result. It extracts the values for 'aspect', 'polarity', 'opinion', and 'category' from each dictionary and appends them as a new row to the DataFrame ``df``.

Ensure that `df` is initialized with the correct columns before running this loop. If you encounter any errors or need further assistance, feel free to ask!



You

```
AttributeError           Traceback (most recent call last)
<ipython-input-22-2d63e5be7e76> in <cell line: 2>()
      2 for i, absa_result in enumerate(absa_results):
      3     for quad in absa_result["Quadruples"]:
----> 4         df = df.append({'aspect': quad['aspect'],
      5                     'polarity': quad['polarity'],
      6                     'opinion': quad['opinion'],

/usr/local/lib/python3.10/dist-packages/pandas/core/generic.py in __getattr__(self, name)
5987     ):
5988     return self[name]
-> 5989     return object.__getattribute__(self, name)
5990
5991 def __setattr__(self, name: str, value) -> None:

```



ChatGPT

It seems there was an error because the DataFrame '`df`' does not have an '`append`' attribute. You can use the '`concat`' function from pandas to concatenate DataFrames. Here's how you can modify the code to use '`concat`':

```
python Copy code
import pandas as pd
# Initialize an empty DataFrame with the desired columns
```

Message ChatGPT

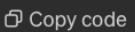
ChatGPT can make mistakes. Check important info.

 You

can we use directly like df.loc[]

 ChatGPT

Yes, you can use the `df.loc[]` method to directly assign values to specific rows and columns in the DataFrame. Here's how you can modify the code to use `df.loc[]`:

```
python 
```

```
import pandas as pd

# Initialize an empty DataFrame with the desired columns
df = pd.DataFrame(columns=['aspect', 'polarity', 'opinion', 'category'])

# Iterate over the rows of the DataFrame and extract the desired information
for absa_result in absa_results:
    for quad in absa_result['Quadruples']:
        row_index = len(df) # Get the current length of the DataFrame as 1
        df.loc[row_index, 'aspect'] = quad['aspect']
        df.loc[row_index, 'polarity'] = quad['polarity']
        df.loc[row_index, 'opinion'] = quad['opinion']
        df.loc[row_index, 'category'] = quad['category']
```

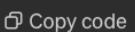
This code initializes an empty DataFrame `df` with the desired columns. Then, it iterates over each element of `absa_results` and assigns the values for 'aspect', 'polarity', 'opinion', and 'category' directly to the corresponding row and column in the DataFrame using `df.loc[]`.

 You

I have an existing df

 ChatGPT

If you have an existing DataFrame `df` and you want to add rows extracted from `absa_results` to it, you can use the `append()` method along with `ignore_index=True` to ensure that the index values are updated accordingly. Here's how you can do it:

```
python 
```

```
import pandas as pd

# Assuming you already have an existing DataFrame named df
# Initialize an empty list to store dictionaries representing rows
rows_to_add = []

# Iterate over the elements of absa_results and extract the desired information
for absa_result in absa_results:
    for quad in absa_result['Quadruples']:
        row = {'aspect': quad['aspect'],
               'polarity': quad['polarity'],
               'opinion': quad['opinion']}.
```

Message ChatGPT 

ChatGPT can make mistakes. Check important info.

```
# Initialize an empty list to store dictionaries representing rows
rows_to_add = []

# Iterate over the elements of absa_results and extract the desired information
for absa_result in absa_results:
    for quad in absa_result['Quadruples']:
        row = {'aspect': quad['aspect'],
               'polarity': quad['polarity'],
               'opinion': quad['opinion'],
               'category': quad['category']}
        rows_to_add.append(row)

# Append the rows to the existing DataFrame
df = df.append(rows_to_add, ignore_index=True)
```

This code iterates over each element of `absa_results`, extracts the desired information for each 'Quadruples' element, and appends it to the `rows_to_add` list as dictionaries representing rows. Then, it appends these rows to the existing DataFrame `df` using the `append()` method with `ignore_index=True` to ensure that the index values are updated properly.



You
earlier I just told you there's no df.append



ChatGPT

Apologies for the oversight. If you don't want to use `df.append`, you can directly use `df.loc` to set the values for each row. Here's how you can do it:

```
python Copy code

# Initialize an empty list to store dictionaries representing rows
rows_to_add = []

# Iterate over the elements of absa_results and extract the desired information
for absa_result in absa_results:
    for quad in absa_result['Quadruples']:
        row = {'aspect': quad['aspect'],
               'polarity': quad['polarity'],
               'opinion': quad['opinion'],
               'category': quad['category']}
        rows_to_add.append(row)

# Iterate over the list of dictionaries and set values in the DataFrame using df.loc
for i, row_data in enumerate(rows_to_add):
    df.loc[len(df)] = row_data
```

This code achieves the same result as before but without using `df.append`. Instead, it directly sets values in the DataFrame `df` using `df.loc`, ensuring that new rows are added with the correct index values.

Message ChatGPT

ChatGPT can make mistakes. Check important info.

```
# within each brand Which aspect appears the most in the reviews?

aspect_counts = test_df.groupby(['brand',
'defined_aspect']).size().reset_index(name='counts')

# For better visualization, we might want to sort the data
aspect_counts = aspect_counts.sort_values(['brand', 'counts'], ascending=[True, False])

plt.figure(figsize=(14, 8))
sns.barplot(data=aspect_counts, x='defined_aspect', y='counts', hue='brand', dodge=True)

# Enhance the plot with labels and a title
plt.title('Most Common Aspects in Reviews by Brand')
plt.xlabel('Aspects')
plt.ylabel('Frequency of Mentions')
plt.xticks(rotation=45) # Rotate the aspect names for better readability

# Adjust the legend
plt.legend(title='Brand', title_fontsize='13', loc='upper right')

# Show the plot
plt.tight_layout()
plt.show() ---- modify this code based on clean_df
```



ChatGPT

To modify your existing plotting code to visualize the most frequently mentioned aspects in reviews by brand based on your `clean_df` dataset, ensure that your DataFrame has the columns `brand` and `defined_aspect` properly set. Here's how you can adjust the code to use `clean_df` correctly, including ensuring the plot is clear and informative:

python

Copy code

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

# Assuming clean_df is your DataFrame and it's already loaded

# Count the occurrences of each aspect within each brand
aspect_counts = clean_df.groupby(['brand', 'defined_aspect']).size().reset_index(name='counts')

# Sort the data to show the most common aspects within each brand
aspect_counts = aspect_counts.sort_values(['brand', 'counts'], ascending=[True, False])
```

```
# within each brand Which aspect appears the most in the reviews?

aspect_counts = test_df.groupby(['brand',
'defined_aspect']).size().reset_index(name='counts')

# For better visualization, we might want to sort the data
aspect_counts = aspect_counts.sort_values(['brand', 'counts'], ascending=[True, False])

plt.figure(figsize=(14, 8))
sns.barplot(data=aspect_counts, x='defined_aspect', y='counts', hue='brand', dodge=True)

# Enhance the plot with labels and a title
plt.title('Most Common Aspects in Reviews by Brand')
plt.xlabel('Aspects')
plt.ylabel('Frequency of Mentions')
plt.xticks(rotation=45) # Rotate the aspect names for better readability

# Adjust the legend
plt.legend(title='Brand', title_fontsize='13', loc='upper right')

# Show the plot
plt.tight_layout()
plt.show() ---- modify this code based on clean_df
```



ChatGPT

To modify your existing plotting code to visualize the most frequently mentioned aspects in reviews by brand based on your `clean_df` dataset, ensure that your DataFrame has the columns `brand` and `defined_aspect` properly set. Here's how you can adjust the code to use `clean_df` correctly, including ensuring the plot is clear and informative:

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

# Assuming clean_df is your DataFrame and it's already loaded

# Count the occurrences of each aspect within each brand
aspect_counts = clean_df.groupby(['brand', 'defined_aspect']).size().reset_index(name='counts')

# Sort the data to show the most common aspects within each brand
aspect_counts = aspect_counts.sort_values(['brand', 'counts'], ascending=[True, False])

# Create the plot
plt.figure(figsize=(14, 8))
sns.barplot(data=aspect_counts, x='defined_aspect', y='counts', hue='brand', dodge=True)

# Enhance the plot with labels and a title
plt.title('Most Common Aspects in Reviews by Brand')
plt.xlabel('Aspects')
plt.ylabel('Frequency of Mentions')
plt.xticks(rotation=45) # Rotate the aspect names for better readability

# Adjust the legend to be more informative
plt.legend(title='Brand', title_fontsize='13', loc='upper right')

# Ensure everything fits without clipping
plt.tight_layout()
```

```

aspect_counts = test_df.groupby(['brand',
                               'defined_aspect']).size().reset_index(name='counts')

# For better visualization, we might want to sort the data
aspect_counts = aspect_counts.sort_values(['brand', 'counts'], ascending=[True, False])

plt.figure(figsize=(14, 8))
sns.barplot(data=aspect_counts, x='defined_aspect', y='counts', hue='brand', dodge=True)

# Enhance the plot with labels and a title
plt.title('Most Common Aspects in Reviews by Brand')
plt.xlabel('Aspects')
plt.ylabel('Frequency of Mentions')
plt.xticks(rotation=45) # Rotate the aspect names for better readability

# Adjust the legend
plt.legend(title='Brand', title_fontsize='13', loc='upper right')

# Show the plot
plt.tight_layout()
plt.show() ---- modify this code based on clean_df
∅

```



ChatGPT

To modify your existing plotting code to visualize the most frequently mentioned aspects in reviews by brand based on your `clean_df` dataset, ensure that your DataFrame has the columns `brand` and `defined_aspect` properly set. Here's how you can adjust the code to use `clean_df` correctly, including ensuring the plot is clear and informative:

```

python
Copy code

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

# Assuming clean_df is your DataFrame and it's already loaded

# Count the occurrences of each aspect within each brand
aspect_counts = clean_df.groupby(['brand', 'defined_aspect']).size().reset_index(name='counts')

# Sort the data to show the most common aspects within each brand
aspect_counts = aspect_counts.sort_values(['brand', 'counts'], ascending=[True, False])

# Create the plot
plt.figure(figsize=(14, 8))
sns.barplot(data=aspect_counts, x='defined_aspect', y='counts', hue='brand', dodge=True)

# Enhance the plot with labels and a title
plt.title('Most Common Aspects in Reviews by Brand')
plt.xlabel('Aspects')
plt.ylabel('Frequency of Mentions')
plt.xticks(rotation=45) # Rotate the aspect names for better readability

# Adjust the legend to be more informative
plt.legend(title='Brand', title_fontsize='13', loc='upper right')

# Ensure everything fits without clipping
plt.tight_layout()

# Display the plot

```

**ChatGPT**

Given your detailed DataFrame named `clean_df`, you can perform a variety of data manipulations and visualizations. If you need to perform any specific tasks like calculating correlations or visualizing the relationship between aspects and ratings or sentiments, you could:

1. Clean the Data:

- Convert the 'price' to numeric after removing commas.
- Handle missing values appropriately.
- Ensure date formats are consistent.

2. Explore Correlations:

- Analyze how different aspects correlate with the overall rating or sentiment scores.

3. Visualize Sentiment Distribution:

- Create visualizations like bar plots or heatmaps to show the distribution of sentiments across different aspects or categories.

4. Sentiment and Aspect Analysis:

- Explore the sentiment polarity in reviews related to different aspects to understand customer preferences or pain points.

Cleaning the 'price' Column

Here's a Python snippet to clean the 'price' column:

```
python
Copy code

import pandas as pd

# Sample DataFrame loading (assuming clean_df is already loaded)
# clean_df = pd.read_csv('path_to_your_csv')

# Convert 'price' to numeric
clean_df['price'] = clean_df['price'].replace(',', '', regex=True).astype(float)
```

Correlation Analysis

If you want to see how different aspects correlate with the overall rating, you could do:

```
python
Copy code

def calculate_correlation(group):
    """ Calculate correlation and handle groups with insufficient data """
    if len(group) > 1: # More than one data point is needed for correlation
        return group['VADER_Score'].corr(group['overall_rating'])
    return None # Return None for groups that cannot have a correlation calculated

# Group by 'brand' and 'defined_aspect' and apply the correlation calculation
correlation_results = clean_df.groupby(['brand', 'defined_aspect']).apply(calculate_c
print(correlation_results)
```

References

- [1] Ambreen Nazir, Y. R., Lianwei Wu, Ling Sun. (2019) 'Issues and Challenges of Aspect-based Sentiment Analysis: A Comprehensive Survey', *IEEE TRANSACTIONS ON AFFECTIVE COMPUTING*.
- [2] D'Aniello, G., Gaeta, M. & La Rocca, I. (2022) 'KnowMIS-ABSA: an overview and a reference model for applications of sentiment analysis and aspect-based sentiment analysis'. *Artif Intell Rev* 55, pp. 5543–5574.
- [3] Kolasani, S. (2023) 'Optimizing Natural Language Processing, Large Language Models (LLMs) for Efficient Customer Service, and hyper-personalization to enable sustainable growth and revenue'. *Transactions on Latest Trends in Artificial Intelligence* 4 (31).
- [4] Liu, J. & Toubia, O. (2018) 'A Semantic Approach for Estimating Consumer Content Preferences from Online Search Queries'. *Marketing Science*, 37(6), pp. 930-952.
- [5] Pontiki, M., Galanis, D., Pavlopoulos, J., Papageorgiou, H., Androutsopoulos, I., & Manandhar, S. (2014) 'SemEval-2014 Task 4: Aspect Based Sentiment Analysis. Proceedings of the 8th international workshop on semantic evaluation'. pp. 27-35.
- [6] Rashid, A., Zeb, M.A., Rashid, A., Anwar, S., Joaquim, F. and Halim, Z. (2020) 'Conceptualization of smartphone usage and feature preferences among various demographics'. *Cluster Computing*, pp. 1855-1873.
- [7] Sayeed, M., Mohan, V. & Anbananthen, K. (2023) 'BERT: A Review of Applications in Sentiment Analysis'. *HighTech and Innovation Journal*, 4, pp. 453-462.
- [8] Scaria, K., Gupta, H., Goyal, S., Sawant, S.A., Mishra, S. and Baral, C. (2023) 'Instructabsa: Instruction learning for aspect based sentiment analysis'. arXiv preprint arXiv:2302.08624.
- [9] T. Ananth kumar, Z. J., P. Kanimozhi, Rajmohan R., Christo Ananth, Sunday Adeola Ajagbe. (2024) 'Machine Learning and Sentiment Analysis: Analysing Customer Feedback'.
- [10] Thomas , D. and Mathur , S. , Data analysis by web scraping using Python . (2019) '3rd International Conference on Electronics, Communication and Aerospace Technology (ICECA)' , pp. 450 – 454 .
- [11] Wan Min, W. & Zulkarnain, N. (2020) 'Comparative Evaluation of Lexicons in Performing Sentiment Analysis', 2, pp. 14-20.

- [12] Yang, H., Zhang, C., & Li, K. (2023) ‘PyABSA: A Modularized Framework for Reproducible Aspect-based Sentiment Analysis’. In Proceedings of the 32nd ACM International Conference on Information and Knowledge Management, pp. 5117–5122.