

# **A Start-up powering Aspect-based Sentiment Analysis using NLP**

## **A PROJECT REPORT – FINAL REVIEW**

*Submitted by*

Abhinav Jaiswal - 20BCE2624

Vanshit Kandoi - 20BCE2667

Sneha Jayshri - 20BCE2673

Gaurang Rastogi - 20BCE2954

**Course Code: 4022**

**Course Title: Natural Language Processing**

Under the guidance of

**Prof. Sharmila Banu K**

**VIT, Vellore.**



**VIT<sup>®</sup>**  

---

**Vellore Institute of Technology**  
(Deemed to be University under section 3 of UGC Act, 1956)

**SCHOOL OF COMPUTER SCIENCE AND ENGINEERING**

**April, 2023**

## GITHUB LINKS:

1. <https://github.com/kevinscaria/InstructABSA>
2. <https://github.com/anmolbansal7/Auto-NLP>
3. <https://github.com/Prajwal10031999/Sentiment-Analysis-ML-Flask-App>
4. <https://github.com/yangheng95/PyABSA>

## Some Research Papers:

1. <https://reader.elsevier.com/reader/sd/pii/S2666449621000529?token=80E7CD4AAEE012EB603D42ACE7057B4CC155D5E5785BF89B11FEA83934EC1F0298A7ECD203B0425214FCF1022CABC7BB&originRegion=eu-west-1&originCreation=20230414182230>
2. <https://dl.acm.org/doi/10.1145/3389035>

## YouTube API

<https://developers.google.com/youtube/>

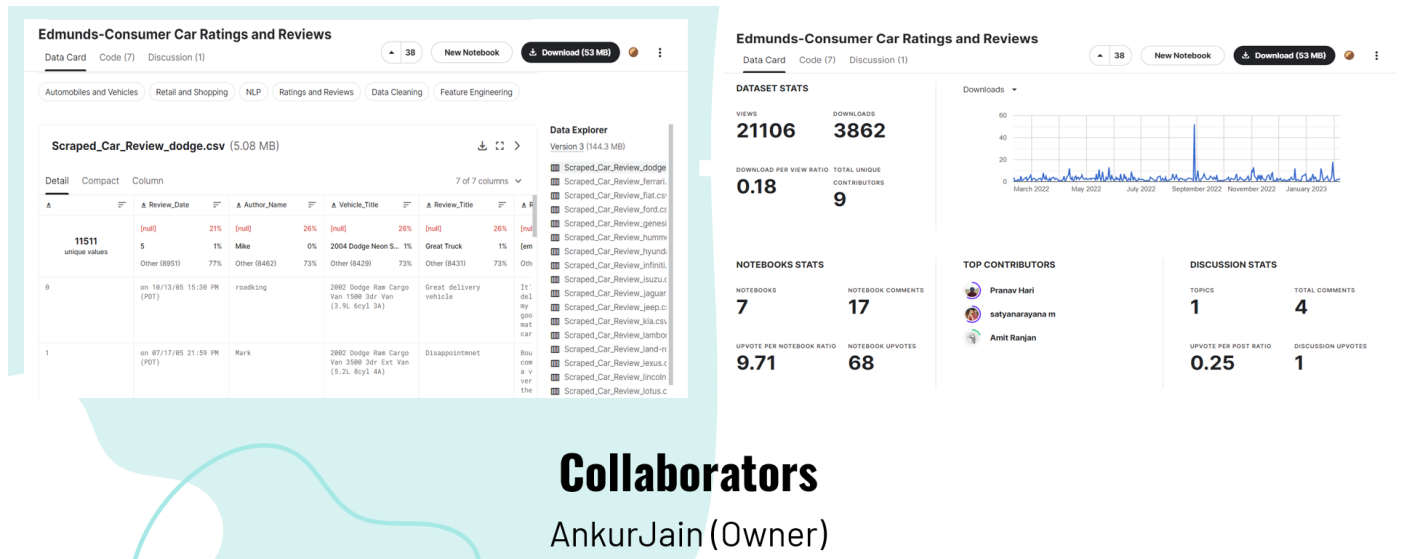
The screenshot shows the YouTube Data API v3 documentation page for the 'comments' resource. The page is titled 'Resource representation' and explains the JSON structure for a 'comments' resource. The JSON structure is as follows:

```
{
  "kind": "youtube#comment",
  "etag": "etag",
  "id": "string",
  "snippet": {
    "authorDisplayName": "string",
    "authorProfileImageUrl": "string",
    "authorChannelUrl": "string",
    "authorChannelId": {
      "value": "string"
    },
    "channelId": "string",
    "videoId": "string",
    "textDisplay": "string",
    "textOriginal": "string",
    "parentId": "string",
    "canRate": "boolean",
    "viewerRating": "string",
    "likeCount": "unsigned integer",
    "moderationStatus": "string",
    "publishedAt": "datetime",
    "updatedAt": "datetime"
  }
}
```

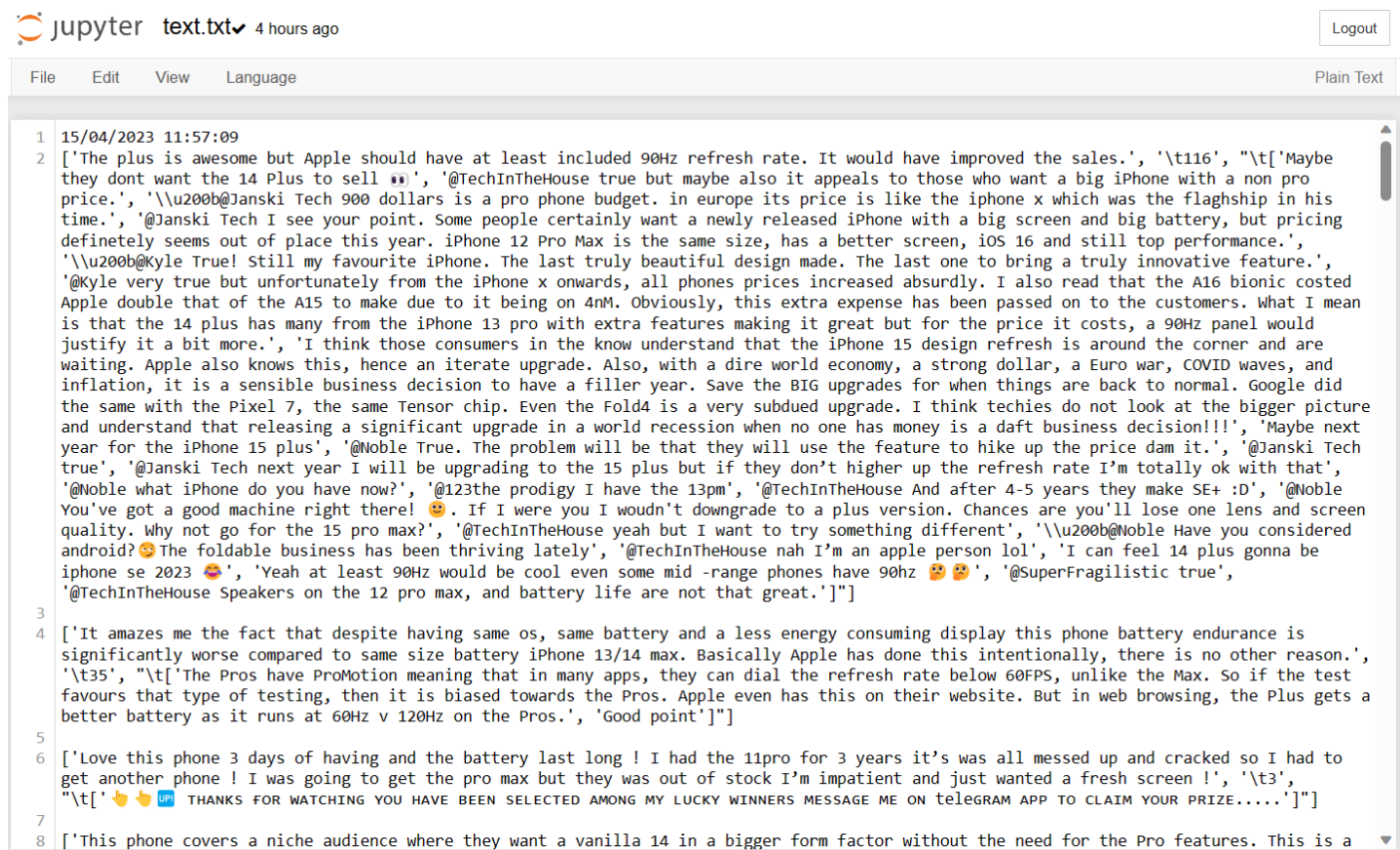
The page also includes a sidebar with navigation links for various YouTube API resources, a search bar, and a 'Recommended for you' section.

# Datasets:

## 1. [Edmunds-Consumer Car Ratings and Reviews | Kaggle](#)



## 2. Created by YouTube comments by ourselves.



# Our Work:

## Aspect Based Sentiment Analysis: Find Available Checkpoints For Current Version

```
In [1]: from pyabsa import available_checkpoints
checkpoint_map = available_checkpoints()

No CUDA GPU found in your device
[2023-04-15 13:28:03] (2.2.3) PyABSA(2.2.3):
[New Feature] Aspect Sentiment Triplet Extraction from v2.1.0 test version (https://github.com/yangheng95/PyABSA/tree/v2/examples-v2/aspect_sentiment_triplet_extraction)
[New Feature] Aspect CategoryOpinion Sentiment Quadruple Extraction from v2.2.0 test version (https://github.com/yangheng95/PyABSA/tree/v2/examples-v2/aspect_opinion_sentiment_category_extraction)
```

If you find any problems, please report them on GitHub. Thanks!  
The v2.x versions are not compatible with Google Colab. Please downgrade to 1.16.27.

[2023-04-15 13:28:03] (2.2.3) Please specify the task code, e.g. from pyabsa import TaskCodeOption

```
C:\Users\vansh\anaconda3\lib\multiprocessing\pool.py:265: ResourceWarning: unclosed running multiprocessing pool <multiprocessing.pool.Pool state=RUN pool_size=1>
  _warn(f"unclosed running multiprocessing pool {self!r}",
ResourceWarning: Enable tracemalloc to get the object allocation traceback
```

## Init a aspect extractor from online checkpoint or local checkpoint

```
In [2]: from pyabsa import ATEPCCheckpointManager

aspect_extractor = ATEPCCheckpointManager.get_aspect_extractor(checkpoint='english',
                                                             auto_device=True # False means load model on CPU
                                                             )

[2023-04-15 13:28:23] (2.2.3) ***** Available ATEPC model checkpoints for Version:2.2.3 (this version) *****
[2023-04-15 13:28:23] (2.2.3) ***** Available ATEPC model checkpoints for Version:2.2.3 (this version) *****
[2023-04-15 13:28:23] (2.2.3) Downloading checkpoint:english
[2023-04-15 13:28:23] (2.2.3) Notice: The pretrained model are used for testing, it is recommended to train the model on your own custom datasets
[2023-04-15 13:28:23] (2.2.3) Checkpoint already downloaded, skip
[2023-04-15 13:28:23] (2.2.3) Load aspect extractor from checkpoints\ATEPC_ENGLISH_CHECKPOINT\fast_lcf_atepc_English_cdw_apcacc_82.36_apcf1_81.89_atef1_75.43
[2023-04-15 13:28:23] (2.2.3) config: checkpoints\ATEPC_ENGLISH_CHECKPOINT\fast_lcf_atepc_English_cdw_apcacc_82.36_apcf1_81.89_atef1_75.43\fast_lcf_atepc.config
[2023-04-15 13:28:23] (2.2.3) state_dict: checkpoints\ATEPC_ENGLISH_CHECKPOINT\fast_lcf_atepc_English_cdw_apcacc_82.36_apcf1_81.89_atef1_75.43\fast_lcf_atepc.state_dict
[2023-04-15 13:28:23] (2.2.3) model: None
[2023-04-15 13:28:23] (2.2.3) tokenizer: checkpoints\ATEPC_ENGLISH_CHECKPOINT\fast_lcf_atepc_English_cdw_apcacc_82.36_apcf1_81.89_atef1_75.43\fast_lcf_atepc.tokenizer
[2023-04-15 13:28:23] (2.2.3) Set Model Device: cpu
[2023-04-15 13:28:23] (2.2.3) Device Name: Unknown
```

Some weights of the model checkpoint at microsoft/deberta-v3-base were not used when initializing DebertaV2Model: ['mask\_predictions.dense.bias', 'mask\_predictions.LayerNorm.weight', 'mask\_predictions.LayerNorm.bias', 'mask\_predictions.classifier.weight', 'lm\_predictions.lm\_head.dense.bias', 'lm\_predictions.lm\_head.bias', 'mask\_predictions.dense.weight', 'lm\_predictions.lm\_head.dense.weight', 'mask\_predictions.classifier.bias', 'lm\_predictions.lm\_head.LayerNorm.bias', 'lm\_predictions.lm\_head.LayerNorm.weight']

- This IS expected if you are initializing DebertaV2Model from the checkpoint of a model trained on another task or with another architecture (e.g. initializing a BertForSequenceClassification model from a BertForPreTraining model).

- This IS NOT expected if you are initializing DebertaV2Model from the checkpoint of a model that you expect to be exactly identical (initializing a BertForSequenceClassification model from a BertForSequenceClassification model).

Special tokens have been added in the vocabulary, make sure the associated word embeddings are fine-tuned or trained.

C:\Users\vansh\anaconda3\lib\site-packages\transformers\convert\_slow\_tokenizer.py:446: UserWarning: The sentencepiece tokenizer that you are converting to a fast tokenizer uses the byte fallback option which is not implemented in the fast tokenizers. In p

# Aspect Term Extract & Sentiment Inference

Jupyter Aspect\_Based\_Sentiment\_Analysis Last Checkpoint: 2 hours ago (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help

Notebook saved

Not Trusted

Python 3 (ipykernel)

Run

```
In [8]: # You can inference from a list of sentences or a DatasetItem from PyABSA
examples = ['Staff was very rude but food was delicious']
inference_source = examples
atepc_result = aspect_extractor.extract_aspect(inference_source=inference_source, #
                                              pred_sentiment=True, # Predict the sentiment of extracted aspect terms
                                              )
```

C:\Users\vansh\anaconda3\lib\site-packages\pyabsa\tasks\AspectTermExtraction\prediction\aspect\_extractor.py:563: UserWarning: Creating a tensor from a list of numpy.ndarrays is extremely slow. Please consider converting the list to a single numpy.ndarray with numpy.array() before converting to a tensor. (Triggered internally at C:\b\abs\_bao0hdcrdh\croot\pytorch\_1675190257512\work\torch\csrc\utils\tensor\_new.cpp:204.)

```
lcf_cdm_vec = torch.tensor(
```

[2023-04-14 23:14:51] (2.2.3) The results of aspect term extraction have been saved in C:\Users\vansh\Desktop\NLP\_Proj\PyABSA-2\Aspect Term Extraction and Polarity Classification.FAST\_LCF\_ATEPC.result.json

[2023-04-14 23:14:51] (2.2.3) Example 0: Staff was very rude but <food:Positive Confidence:0.9994> was delicious

C:\Users\vansh\anaconda3\lib\site-packages\pyabsa\tasks\AspectTermExtraction\prediction\aspect\_extractor.py:644: UserWarning: Implicit dimension choice for softmax has been deprecated. Change the call to include dim=X as an argument.

```
float(x) for x in F.softmax(i_apc_logits).cpu().numpy().tolist()
```

```
In [9]: atepc_result
```

```
Out[9]: [{'sentence': 'Staff was very rude but food was delicious',
          'IOB': ['B-ASP', 'O', 'O', 'O', 'O', 'B-ASP', 'O', 'O'],
          'tokens': ['Staff',
                    'was',
                    'very',
                    'rude',
                    'but',
                    'food',
                    'was',
                    'delicious'],
          'aspect': ['Staff', 'food'],
          'position': [[1], [6]],
          'sentiment': ['Negative', 'Positive'],
          'probs': [[0.9969356060028076, 0.002819894580170512, 0.0002444844867568463],
                   [0.00032588167232461274, 0.0002935877101151645, 0.999380469322046]],
          'confidence': [0.9969, 0.9994]]]
```

Jupyter Aspect\_Based\_Sentiment\_Analysis Last Checkpoint: 2 hours ago (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help

Not Trusted

Python 3 (ipykernel)

Code

```
In [9]: # You can inference from a list of sentences or a DatasetItem from PyABSA
examples = ['Its an fine okayish phone']
inference_source = examples
atepc_result = aspect_extractor.extract_aspect(inference_source=inference_source, #
                                              pred_sentiment=True, # Predict the sentiment of extracted aspect terms
                                              )

atepc_result
```

[2023-04-15 11:20:24] (2.2.3) The results of aspect term extraction have been saved in C:\Users\vansh\Desktop\NLP\_Proj\PyABSA-2\Aspect Term Extraction and Polarity Classification.FAST\_LCF\_ATEPC.result.json

[2023-04-15 11:20:24] (2.2.3) Example 0: Its an fine okayish phone

```
Out[9]: [{'sentence': 'Its an fine okayish phone',
          'IOB': ['O', 'O', 'O', 'O', 'B-ASP'],
          'tokens': ['Its', 'an', 'fine', 'okayish', 'phone'],
          'aspect': ['phone'],
          'position': [[5]],
          'sentiment': ['Neutral'],
          'probs': [[0.023734791204333305, 0.9609071612358093, 0.015358027070760727]],
          'confidence': [0.9609]]]
```

```
In [51]: # You can inference from a list of sentences or a DatasetItem from PyABSA
examples = [['The display of this laptop sometimes flicker, but it has a genuine price'], ['Its an fine okayish phone']]
atepc_result = []
for inference_source in examples:
    atepc_result.append(aspect_extractor.extract_aspect(inference_source=inference_source, #
                                                       pred_sentiment=True, # Predict the sentiment of extracted aspect terms
                                                       )
                      )

atepc_result
```

[2023-04-15 11:59:08] (2.2.3) The results of aspect term extraction have been saved in C:\Users\vansh\Desktop\NLP\_Proj\PyABSA-2\Aspect Term Extraction and Polarity Classification.FAST\_LCF\_ATEPC.result.json

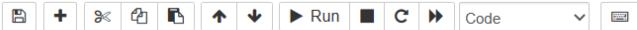
[2023-04-15 11:59:08] (2.2.3) Example 0: The <display:Negative Confidence:0.9985> of this laptop sometimes flicker , but it has a genuine price

[2023-04-15 12:02:21] (2.2.3) The results of aspect term extraction have been saved in C:\Users\vansh\Desktop\NLP\_Proj\PyABSA-2\Aspect Term Extraction and Polarity Classification.FAST\_LCF\_ATEPC.result.json

[2023-04-15 12:02:21] (2.2.3) Example 0: Its an fine okayish phone

```
Out[51]: [[{'sentence': 'The display of this laptop sometimes flicker , but it has a genuine price',
```





```
In [50]: import io
with io.open('text.txt', 'r', encoding='utf8') as f:
    text = f.read()
    text = text[text.find('\n') + 1:]
    text = text.split('\n\n')

    for i in text:
        i = i.strip('[]').split(', ')
        print(i[0])
```

'The plus is awesome but Apple should have at least included 90Hz refresh rate. It would have improved the sales.'

'It amazes me the fact that despite having same os

'Love this phone 3 days of having and the battery last long ! I had the 11pro for 3 years it's was all messed up and cracked so I had to get another phone ! I was going to get the pro max but they was out of stock I'm impatient and just wanted a fres h screen !'

'This phone covers a niche audience where they want a vanilla 14 in a bigger form factor without the need for the Pro feature s. This is a better attempt than the Mini phones because more and more people prioritize battery life and bigger screens.'

'I have this phone and it's a great great phone n the gen 8 watch has much more options paired with the 14 that the 13 don't have n iOS 16 is strong'

'The iPhone 14 has improved audio out from the lightning socket for wired IEMs (I have lighting wired IEMs). Do you know if t he iPhone 14 plus has the same improved audio out conaore to the iPhone 13?'

'Love it

'The iPhone 14 Plus is even lighter than the iPhone XS Max! 🥰'

'It's the exact same as the 13 pro max minus 1 camera and the 120hz screen. I would get which ever one is cheaper. I have the 14pro max. I shouldn't have upgraded from the 13pro max not worth it. The 13pro max is a beast.'

'I have never bought pro models in my life but this time when i upgrade from xr

'It's not a bad phone but paying literally flagship prices for a 60Hz display and last year's chipset is not wise. Go for the Pro Max if you want something big

'Basically if you want the 14 plus go with the 13 pro max,nearly same price plus 120 hertz and longer battery life (according

```
In [86]: def ABSA(examples):
    with io.open('ABSA.txt', 'a', encoding='utf8') as f:
        inference_source = [examples]
        atepc_result = aspect_extractor.extract_aspect(inference_source=inference_source, #
            pred_sentiment=True, # Predict the sentiment of extracted aspect terms
        )
        f.write(str(atepc_result) + "\n\n")
    atepc_result
```

```
In [ ]: import pickle
#define an empty dictionary
l = []

# Load the pickle contents
with open ('comments1.pkl','rb') as pick:
    l.append(pickle.load(pick))
# print(l)
for i in l[0]:
    print(i[0], end='\n\n')
    print(ABSA(i[0]))
```

The plus is awesome but Apple should have at least included 90Hz refresh rate. It would have improved the sales.

[2023-04-15 12:57:51] (2.2.3) The results of aspect term extraction have been saved in C:\Users\vansh\Desktop\NLP\_Proj\PyABSA-2\Aspect Term Extraction and Polarity Classification.FAST\_LCF\_ATEPC.result.json

[2023-04-15 12:57:51] (2.2.3) Example 0: The plus is awesome but Apple should have at least included 90Hz <refresh:Negative Confidence:0.8074> rate . It would have improved the sales .

It amazes me the fact that despite having same os, same battery and a less energy consuming display this phone battery endurance is significantly worse compared to same size battery iPhone 13/14 max. Basically Apple has done this intentionally, there is no other reason.

[2023-04-15 13:00:18] (2.2.3) The results of aspect term extraction have been saved in C:\Users\vansh\Desktop\NLP\_Proj\PyABSA-2\Aspect Term Extraction and Polarity Classification.FAST\_LCF\_ATEPC.result.json

[2023-04-15 13:00:18] (2.2.3) Example 0: It amazes me the fact that despite having same <os:Positive Confidence:0.9865> , same <battery:Positive Confidence:0.9806> and a less energy consuming <display:Positive Confidence:0.9202> this phone <battery endurance:Negative Confidence:0.9648> is significantly worse compared to same size battery iPhone 13 / 14 max . Basically Apple has done this intentionally , there is no other reason .

Love this phone 3 days of having and the battery last long ! I had the 11pro for 3 years it's was all messed up and cracked so I had to get another phone ! I was going to get the pro max but they was out of stock I'm impatient and just wanted a fresh screen !

It amazes me the fact that despite having same os, same battery and a less energy consuming display this phone battery endurance is significantly worse compared to same size battery iPhone 13/14 max. Basically Apple has done this intentionally, there is no other reason.

Love this phone 3 days of having and the battery last long ! I had the 11pro for 3 years it's was all messed up and cracked so I had to get another phone ! I was going to get the pro max but they was out of stock I'm impatient and just wanted a fresh screen !

This is a better attempt than the Mini phones because more and more people prioritize battery life and bigger screens.

and more people prioritize <battery life:Positive Confidence:0.957> and bigger <screens:Positive Confidence:0.9793> .  
I have this phone and it's a great great phone n the gen 8 watch has much more options paired with the 14 that the 13 don't have n iOS 16 is strong

The iPhone 14 has improved audio out from the lightning socket for wired IEMs (I have lightning wired IEMs). Do you know if the iPhone 14 plus has the same improved audio out conaore to the iPhone 13?

[Logout](#)

JSON

```
[{"sentence": "The iPhone 14 has improved audio out from the lightning socket for wired IEMs ( I have lighting wired IEMs ) . Do you know if  
the iPhone 14 plus has the same improved audio out conoare to the iPhone 13 ?", "IOB": ["O", "O", "O", "O", "O", "B-ASP", "O", "O", "O", "O",  
"O", "O", "O", "O", "O", "O", "O", "O", "O", "O", "O", "O", "O", "O", "O", "O", "O", "O", "O", "O", "O", "O", "O", "O", "O", "O", "O", "O", "O", "O", "O", "O", "O",  
"O", "O", "O", "O"], "tokens": ["The", "iPhone", "14", "has", "improved", "audio", "out", "from", "the", "lightning", "socket", "for", "  
wired", "IEMs", "(", "I", "have", "lighting", "wired", "IEMs", ")". ".", "Do", "you", "know", "if", "the", "iphone", "14", "plus", "has",  
"the", "same", "improved", "audio", "out", "conoare", "to", "the", "iphone", "13", "?"], "aspect": ["audio"], "position": [[6]], "sentiment":  
["Positive"], "probs": [[0.0031521590426564217, 0.0007129677105695009, 0.996134877204895]], "confidence": [0.9961]]}
```

[illegible]

## Using VADER Sentiment Analysis:

```
In [10]: import nltk
nltk.download('punkt')
nltk.download('stopwords')
nltk.download('vader_lexicon')

from nltk.sentiment.vader import SentimentIntensityAnalyzer
from nltk.tokenize import word_tokenize
from nltk.corpus import stopwords

# Create an instance of the SentimentIntensityAnalyzer
sia = SentimentIntensityAnalyzer()

# Define a function to clean the text and tokenize it
def preprocess_text(text):
    # Convert text to lowercase
    text = text.lower()
    # Remove stop words
    stop_words = set(stopwords.words('english'))
    words = word_tokenize(text)
    words = [w for w in words if not w in stop_words]
    # Join the words back into a string
    text = ' '.join(words)
    return text

# Define a function to perform sentiment analysis on a given text
def analyze_sentiment(text):
    # Preprocess the text
    text = preprocess_text(text)
    # Analyze the sentiment using the SentimentIntensityAnalyzer
    sentiment = sia.polarity_scores(text)
    return sentiment
```

```
In [20]: import pickle
#define an empty dictionary
l = []

# Load the pickle contents
with open('comments1.pkl','rb') as pick:
    l.append(pickle.load(pick))
# print(l)
for i in l[0]:
    print(i[0], end='\n')
    print(analyze_sentiment(i[0]), end = '\n\n')
```

The plus is awesome but Apple should have at least included 90Hz refresh rate. It would have improved the sales.  
{'neg': 0.0, 'neu': 0.556, 'pos': 0.444, 'compound': 0.802}

It amazes me the fact that despite having same os, same battery and a less energy consuming display this phone battery endurance is significantly worse compared to same size battery iPhone 13/14 max. Basically Apple has done this intentionally, there is no other reason.  
{'neg': 0.103, 'neu': 0.731, 'pos': 0.166, 'compound': 0.228}

Love this phone 3 days of having and the battery last long ! I had the 11pro for 3 years it's was all messed up and cracked so I had to get another phone ! I was going to get the pro max but they was out of stock I'm impatient and just wanted a fresh screen !  
{'neg': 0.153, 'neu': 0.6, 'pos': 0.246, 'compound': 0.5826}

This phone covers a niche audience where they want a vanilla 14 in a bigger form factor without the need for the Pro features. This is a better attempt than the Mini phones because more and more people prioritize battery life and bigger screens.  
{'neg': 0.0, 'neu': 0.84, 'pos': 0.16, 'compound': 0.4939}

I have this phone and it's a great great phone n the gen 8 watch has much more options paired with the 14 that the 13 don't have n iOS 16 is strong  
{'neg': 0.0, 'neu': 0.799, 'pos': 0.201, 'compound': 0.691}



# YouTube Extractor:

jupyter YTCommentExtractor Last Checkpoint: 3 hours ago (unsaved changes)



Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted

Python 3 (ipykernel)

Run Code

```
In [1]: #!pip install google-api-python-client
from googleapiclient.discovery import build
from datetime import datetime
# To remove HTML Tags
from bs4 import BeautifulSoup

video_id = "HDkd_n01tvA"
api_key = 'AIzaSyDx49S-3eGxX4f1tXmPQjrJMZNueaiSk0I'

# recursive function to get all replies in a comment thread
def get_replies(comment_id, token):
    replies_response = yt_object.comments().list(part = 'snippet', maxResults = 100, parentId = comment_id, pageToken = token).execute()

    for reply in replies_response['items']:
        all_comments.append(reply['snippet']['textDisplay'])

    if replies_response.get("nextPageToken"):
        return get_replies(comment_id, replies_response['nextPageToken'])
    else:
        return []

# recursive function to get all comments
def get_comments(youtube, video_id, next_view_token):
    global all_comments

    # check for token
    if len(next_view_token.strip()) == 0:
        all_comments = []

    if next_view_token == '':
        # get the initial response
        comment_list = youtube.commentThreads().list(part = 'snippet', maxResults = 100, videoId = video_id, order = 'relevance')
    else:
        # get the next page response
        comment_list = youtube.commentThreads().list(part = 'snippet', maxResults = 100, videoId = video_id, order='relevance', pageToken = next_view_token)

    # loop through all top level comments
    for comment in comment_list['items']:
        # add comment to list
        all_comments.append([comment['snippet']['topLevelComment']['snippet']['textDisplay'])
        # add Number Of Likes to list
        all_comments[-1].append([comment['snippet']['topLevelComment']['snippet']['likeCount'])
        # get number of replies
        reply_count = comment['snippet']['totalReplyCount']
        all_replies = []
        # if replies greater than 0
        if reply_count > 0:
            # get first 100 replies
            replies_list = youtube.comments().list(part='snippet', maxResults=100, parentId=comment['id']).execute()
            for reply in replies_list['items']:
                # add reply to list
                all_replies.append(reply['snippet']['textDisplay'])

            # check for more replies
            while "nextPageToken" in replies_list:
                token_reply = replies_list['nextPageToken']
                # get next set of 1000 replies
                replies_list = youtube.comments().list(part = 'snippet', maxResults = 1000, parentId = comment['id'], pageToken = token_reply)
                for reply in replies_list['items']:
                    # add reply to list
                    s = reply['snippet']['textDisplay']
                    all_replies.append(s)
```

```

        # add all replies to the comment
        all_comments[-1].append(all_replies[:::-1])

    if "nextPageToken" in comment_list:
        return get_comments(youtube, video_id, comment_list['nextPageToken'])
    else:
        return []

all_comments = []

# build a youtube object using our api key
yt_object = build('youtube', 'v3', developerKey=api_key)

# get all comments and replies
comments = get_comments(yt_object, video_id, '')

# Store all the comments into pickle file
import pickle
with open('comments.pkl', 'wb') as file:
    pickle.dump(all_comments, file)

```

```

In [2]: s = []
with open('comments.pkl', 'rb') as file:
    s.append(pickle.load(file))

```

```

In [9]: for i in s[0]:
        print(i)

```

['The plus is awesome but Apple should have at least included 90Hz refresh rate. It would have improved the sales.', 116, ['May be they dont want the 14 Plus to sell 🙄', '@TechInTheHouse true but maybe also it appeals to those who want a big iPhone with a non pro price.', '\u200b@Janski Tech 900 dollars is a pro phone budget. in europe its price is like the iphone x which was th e flagship in his time.', '@Janski Tech I see your point. Some people certainly want a newly released iPhone with a big screen and big battery, but pricing definately seems out of place this year. iPhone 12 Pro Max is the same size, has a better screen, iOS 16 and still top performance.', '\u200b@Kyle True! Still my favourite iPhone. The last truly beautiful design made. The las t one to bring a truly innovative feature.', '@Kyle very true but unfortunately from the iPhone x onwards, all phones prices in creased absurdly. I also read that the A16 bionic costed Apple double that of the A15 to make due to it being on 4nm. Obviousl y, this extra expense has been passed on to the customers. <br><br>What I mean is that the 14 plus has many from the iPhone 13 pro with extra features making it great but for the price it costs, a 90Hz panel would justify it a bit more.', 'I think those consumers in the know understand that the iPhone 15 design refresh is around the corner and are waiting. Apple also knows this, hence an iterate upgrade. Also, with a dire world economy, a strong dollar, a Euro war, COVID waves, and inflation, it is a sen sible business decision to have a filler year. Save the BIG upgrades for when things are back to normal. Google did the same wi th the Pixel 7, the same Tensor chip. Even the Fold4 is a very subdued upgrade. I think techies do not look at the bigger pictu re and understand that releasing a significant upgrade in a world recession when no one has money is a daft business decisio n!!!', 'Maybe next year for the iPhone 15 plus', '@Noble True. The problem will be that they will use the feature to hike up th e price dam it.', '@Janski Tech true', '@Janski Tech next year I will be upgrading to the 15 plus but if they don't higher up t he refresh rate I'm totally ok with that', '@Noble what iPhone do you have now?', '@123the prodigy I have the 13pm', '@TechInTh eHouse And after 4-5 years they make SE+ :D', '@Noble You&#39;ve got a good machine right there! 😊. If I were you I woudn&#39; t downgrade to a plus version. Chances are you&#39;ll lose one lens and screen quality. Why not go for the 15 pro max?', '@Tech InTheHouse yeah but I want to try something different', '\u200b@Noble Have you considered android? 😊<br>The foldable business h as been thriving lately', '@TechInTheHouse nah I'm an apple person lol', 'I can feel 14 plus gonna be iphone se 2023 🥰', 'Yeah at least 90Hz would be cool even some mid -range phones have 90hz 🥰🥰', '@SuperFragilistic true', '@TechInTheHouse Speakers o n the 12 pro max, and battery life are not that great.']]

['It amazes me the fact that despite having same os, same battery and a less energy consuming display this phone battery endura nce is significantly worse compared to same size battery iPhone 13/14 max. Basically Apple has done this intentionally, there i s no other reason.', 35, ['The Pros have ProMotion meaning that in many apps, they can dial the refresh rate below 60FPS, unlik e the Max. So if the test favours that type of testing, then it is biased towards the Pros. Apple even has this on their websit e. But in web browsing, the Plus gets a better battery as it runs at 60Hz <del>v</del> 120Hz on the Pros.', 'Good point']]

['Love this phone 3 days of having and the battery last long ! I had the 11pro for 3 years it's was all messed up and cracked s o I had to get another phone ! I was going to get the pro max but they was out of stock I'm impatient and just wanted a fresh s creen !', 3, ['👉<br>👉 THANKS FOR WATCHING YOU HAVE BEEN SELECTED AMONG MY LUCKY WINNERS MESSAGE ME ON TELEGRAM APP TO CLAI M YOUR PRIZE.....']]

['This phone covers a niche audience where they want a vanilla 14 in a bigger form factor without the need for the Pro feature s. This is a better attempt than the Mini phones because more and more people prioritize battery life and bigger screens.', 51, ['Yes that&#39;s me', 'It could've been better with at least a 90Hz Screen. But again, the price probably denies my suggestio n']]

## ERRORS:

### 1. While Extraction:

#### a. Instagram:

```
In [15]: import requests

# make a GET request to the Instagram page
response = requests.get("https://www.instagram.com/rvcjindia")

# check if the request was successful
if response.status_code == 200:
    # extract the content of the page
    content = response.content

    # search for the data you want to scrape, for example the user's followers count
    start = content.find("\\"edge_followed_by\\":{\\\"count\\\":") + 27
    end = content.find("}\\\"followed_by_viewer\\\"")
    followers = content[start:end]

    # print the followers count
    print("Followers:", followers)
else:
    print("Request failed with status code", response.status_code)
```

```
-----
TimeoutError                                Traceback (most recent call last)
~\anaconda3\lib\site-packages\urllib3\connection.py in _new_conn(self)
    173     try:
--> 174         conn = connection.create_connection(
    175             (self._dns_host, self.port), self.timeout, **extra_kw

~\anaconda3\lib\site-packages\urllib3\util\connection.py in create_connection(address, timeout, source_address, socket_options)
    95     if err is not None:
--> 96         raise err
```

```
~\anaconda3\lib\site-packages\requests\sessions.py in send(self, request, **kwargs)
    653
    654     # Send the request
--> 655     r = adapter.send(request, **kwargs)
    656
    657     # Total elapsed time of the request (approximately)

~\anaconda3\lib\site-packages\requests\adapters.py in send(self, request, stream, timeout, verify, cert, proxies)
    508
    509     if isinstance(e.reason, _ProxyError):
--> 510         raise ProxyError(e, request=request)
    511
    512     if isinstance(e.reason, _SSLError):

ProxyError: HTTPSConnectionPool(host='www.instagram.com', port=443): Max retries exceeded with url: /rvcjindia (Caused by ProxyError('Cannot connect to proxy.', NewConnectionError('<urllib3.connection.HTTPSConnection object at 0x00000211AA9484F0>: Failed to establish a new connection: [WinError 10060] A connection attempt failed because the connected party did not properly respond after a period of time, or established connection failed because connected host has failed to respond')))
```

## b. Twitter:

In [11]: `import twint`

```
c = twint.Config()

c.Search = ['Taylor Swift']      # topic
c.Limit = 500                    # number of Tweets to scrape
c.Store_csv = True               # store tweets in a csv file
c.Output = "taylor_swift_tweets.csv"  # path to csv file

twint.run.Search(c)
```

```
-----
RuntimeError                                Traceback (most recent call last)
~\AppData\Local\Temp\ipykernel_11380\3337082665.py in <module>
      8 c.Output = "taylor_swift_tweets.csv"      # path to csv file
      9
--> 10 twint.run.Search(c)

~\anaconda3\lib\site-packages\twint\run.py in Search(config, callback)
    325     config.Profile = False
    326     config.Profile full = False
--> 327     run(config, callback)
    328     if config.Pandas_auto:
    329         storage.panda._autoget("tweet")

~\anaconda3\lib\site-packages\twint\run.py in run(config, callback)
    224         raise
    225
--> 226     get_event_loop().run_until_complete(Twint(config).main(callback))
    227     get_event_loop().run_until_complete(main(config, callback))
    228 def Favorites(config):

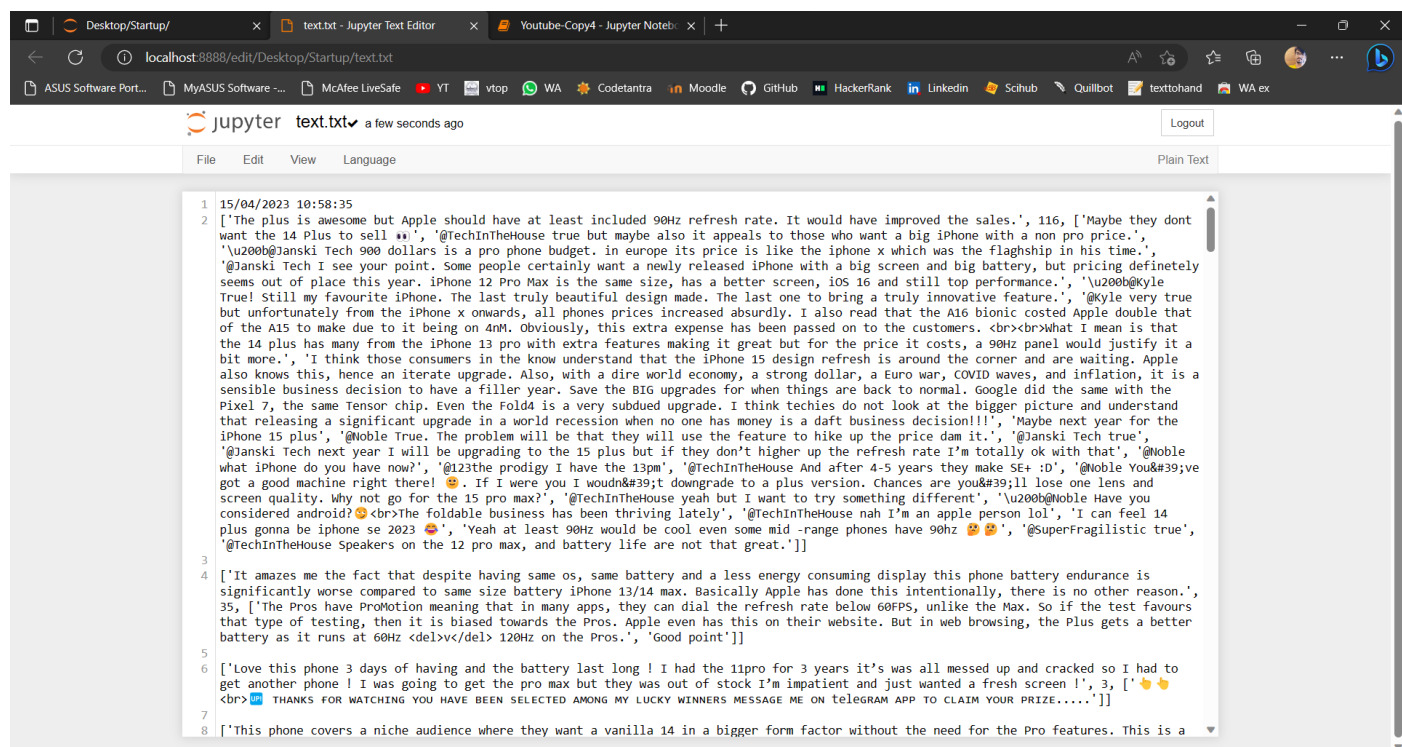
~\anaconda3\lib\asyncio\base_events.py in run_until_complete(self, future)
    590     """
    591     self._check_closed()
--> 592     self._check_running()
    593
    594     new_task = not futures.isfuture(future)

~\anaconda3\lib\asyncio\base_events.py in _check_running(self)
    550     def _check_running(self):
    551         if self.is_running():
--> 552             raise RuntimeError('This event loop is already running')
    553         if events._get_running_loop() is not None:
    554             raise RuntimeError(

RuntimeError: This event loop is already running
```

## c. Youtube

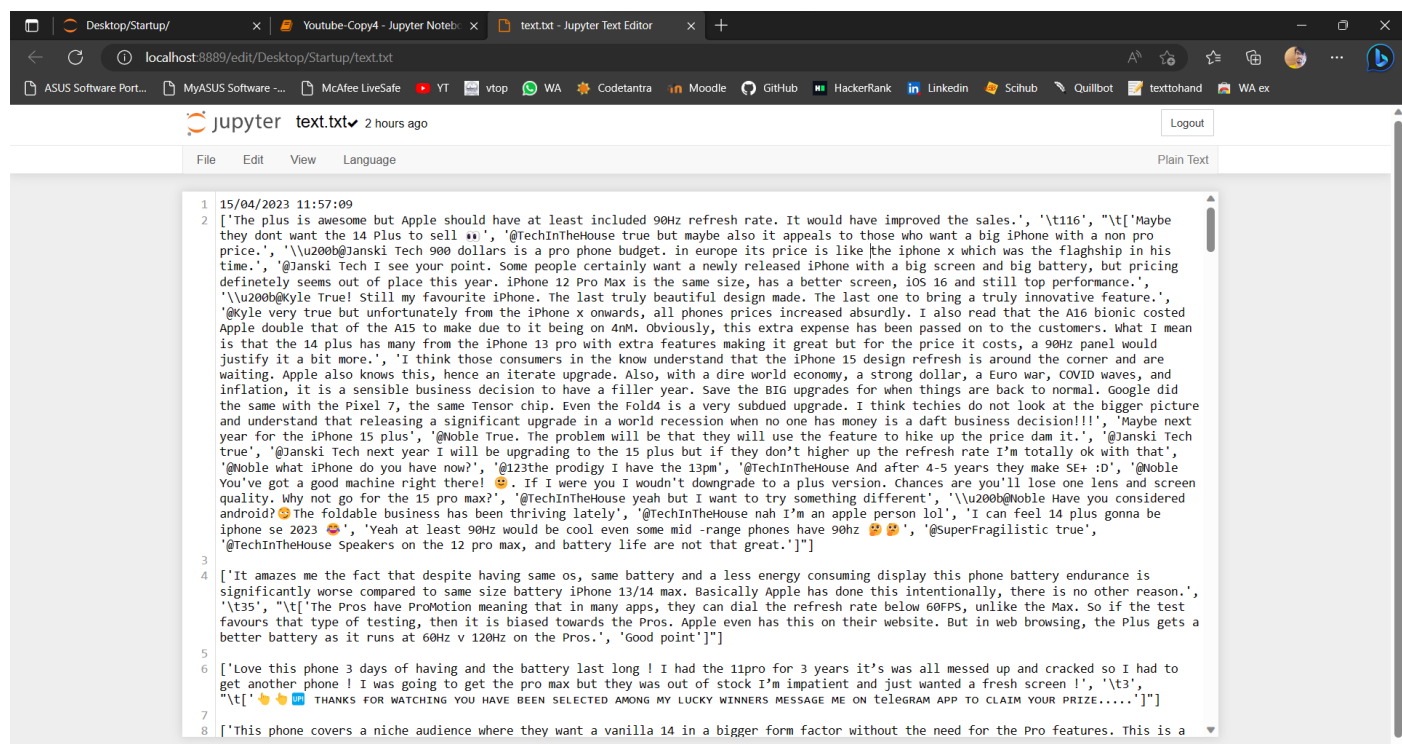
### HTML Tags are there in the text



The screenshot shows a Jupyter Notebook interface with a text file named 'text.txt' open. The text file contains a long paragraph of text, which is a mix of English and HTML tags. The text is as follows:

```
1 15/04/2023 10:58:35
2 ['The plus is awesome but Apple should have at least included 90Hz refresh rate. It would have improved the sales.', '116', ['Maybe they dont
want the 14 Plus to sell *!'], '@TechInTheHouse true but maybe also it appeals to those who want a big iPhone with a non pro price.',
'\u200b@Janski Tech 900 dollars is a pro phone budget. in europe its price is like the iPhone X which was the flagship in his time.',
'\u200b@Janski Tech I see your point. Some people certainly want a newly released iPhone with a big screen and big battery, but pricing definitely
seems out of place this year. iPhone 12 Pro Max is the same size, has a better screen, iOS 16 and still top performance.', '\u200b@Kylie
True! Still my favourite iPhone. The last truly beautiful design made. The last one to bring a truly innovative feature.', '@Kylie very true but
unfortunately from the iPhone X onwards, all phones prices increased absurdly. I also read that the A16 bionic costed Apple double that
of the A15 to make due to it being on 4nm. Obviously, this extra expense has been passed on to the customers. <br><br>What I mean is that
the 14 plus has many from the iPhone 13 pro with extra features making it great but for the price it costs, a 90Hz panel would justify it a
bit more.', 'I think those consumers in the know understand that the iPhone 15 design refresh is around the corner and are waiting. Apple
also knows this, hence an iterate upgrade. Also, with a dire world economy, a strong dollar, a Euro war, COVID waves, and inflation, it is a
sensible business decision to have a filler year. Save the BIG upgrades for when things are back to normal. Google did the same with the
Pixel 7, the same Tensor chip. Even the Fold4 is a very subdued upgrade. I think techies do not look at the bigger picture and understand
that releasing a significant upgrade in a world recession when no one has money is a daft business decision!!!', 'Maybe next year for the
iPhone 15 plus', '@Noble True. The problem will be that they will use the feature to hike up the price dam it.', '@Janski Tech true',
'\u200b@Janski Tech next year I will be upgrading to the 15 plus but if they don't higher up the refresh rate I'm totally ok with that', '@Noble
what iPhone do you have now?', '@123the prodigy I have the 13pm', '@TechInTheHouse And after 4-5 years they make SE+ :D', '@Noble
You've got a good machine right there! 🥳. If I were you I wouldn't downgrade to a plus version. Chances are you'll lose one lens and screen
quality. Why not go for the 15 pro max?', '@TechInTheHouse yeah but I want to try something different', '\u200b@Noble Have you considered
android? 🤖<br>The foldable business has been thriving lately', '@TechInTheHouse nah I'm an apple person lol', 'I can feel 14
plus gonna be iPhone SE 2023 🥳', 'Yeah at least 90Hz would be cool even some mid-range phones have 90Hz 🥳🥳', '@SuperFragilistic true',
'\u200b@TechInTheHouse Speakers on the 12 pro max, and battery life are not that great.'])
3
4 ['It amazes me the fact that despite having same OS, same battery and a less energy consuming display this phone battery endurance is
significantly worse compared to same size battery iPhone 13/14 max. Basically Apple has done this intentionally, there is no other reason.',
'35', ['The Pros have ProMotion meaning that in many apps, they can dial the refresh rate below 60FPS, unlike the Max. So if the test favours
that type of testing, then it is biased towards the Pros. Apple even has this on their website. But in web browsing, the Plus gets a better
battery as it runs at 60Hz <del>v</del> 120Hz on the Pros.', 'Good point']]
5
6 ['Love this phone 3 days of having and the battery last long! I had the 11pro for 3 years it's was all messed up and cracked so I had to
get another phone! I was going to get the pro max but they was out of stock I'm impatient and just wanted a fresh screen!', '3', ['🥳🥳
<br>👉 THANKS FOR WATCHING YOU HAVE BEEN SELECTED AMONG MY LUCKY WINNERS MESSAGE ME ON TELEGRAM APP TO CLAIM YOUR PRIZE.....']]
7
8 ['This phone covers a niche audience where they want a vanilla 14 in a bigger form factor without the need for the Pro features. This is a
```

### Issue Solved



The screenshot shows the same Jupyter Notebook interface with the 'text.txt' file open. The text file contains the same long paragraph of text, which is a mix of English and HTML tags. The text is as follows:

```
1 15/04/2023 11:57:09
2 ['The plus is awesome but Apple should have at least included 90Hz refresh rate. It would have improved the sales.', '\t116', "\t['Maybe
they dont want the 14 Plus to sell *!'], '@TechInTheHouse true but maybe also it appeals to those who want a big iPhone with a non pro
price.', '\u200b@Janski Tech 900 dollars is a pro phone budget. in europe its price is like the iPhone X which was the flagship in his
time.', '@Janski Tech I see your point. Some people certainly want a newly released iPhone with a big screen and big battery, but pricing
definitely seems out of place this year. iPhone 12 Pro Max is the same size, has a better screen, iOS 16 and still top performance.',
'\u200b@Kylie True! Still my favourite iPhone. The last truly beautiful design made. The last one to bring a truly innovative feature.',
'\u200b@Kylie very true but unfortunately from the iPhone X onwards, all phones prices increased absurdly. I also read that the A16 bionic costed
Apple double that of the A15 to make due to it being on 4nm. Obviously, this extra expense has been passed on to the customers. What I mean
is that the 14 plus has many from the iPhone 13 pro with extra features making it great but for the price it costs, a 90Hz panel would
justify it a bit more.', 'I think those consumers in the know understand that the iPhone 15 design refresh is around the corner and are
waiting. Apple also knows this, hence an iterate upgrade. Also, with a dire world economy, a strong dollar, a Euro war, COVID waves, and
inflation, it is a sensible business decision to have a filler year. Save the BIG upgrades for when things are back to normal. Google did
the same with the Pixel 7, the same Tensor chip. Even the Fold4 is a very subdued upgrade. I think techies do not look at the bigger picture
and understand that releasing a significant upgrade in a world recession when no one has money is a daft business decision!!!', 'Maybe next
year for the iPhone 15 plus', '@Noble True. The problem will be that they will use the feature to hike up the price dam it.', '@Janski Tech
true', '@Janski Tech next year I will be upgrading to the 15 plus but if they don't higher up the refresh rate I'm totally ok with that',
'\u200b@Noble what iPhone do you have now?', '@123the prodigy I have the 13pm', '@TechInTheHouse And after 4-5 years they make SE+ :D', '@Noble
You've got a good machine right there! 🥳. If I were you I wouldn't downgrade to a plus version. Chances are you'll lose one lens and screen
quality. Why not go for the 15 pro max?', '@TechInTheHouse yeah but I want to try something different', '\u200b@Noble Have you considered
android? 🤖<br>The foldable business has been thriving lately', '@TechInTheHouse nah I'm an apple person lol', 'I can feel 14 plus gonna be
iPhone SE 2023 🥳', 'Yeah at least 90Hz would be cool even some mid-range phones have 90Hz 🥳🥳', '@SuperFragilistic true',
'\u200b@TechInTheHouse Speakers on the 12 pro max, and battery life are not that great.'])
3
4 ['It amazes me the fact that despite having same OS, same battery and a less energy consuming display this phone battery endurance is
significantly worse compared to same size battery iPhone 13/14 max. Basically Apple has done this intentionally, there is no other reason.',
'\t35', "\t['The Pros have ProMotion meaning that in many apps, they can dial the refresh rate below 60FPS, unlike the Max. So if the test
favours that type of testing, then it is biased towards the Pros. Apple even has this on their website. But in web browsing, the Plus gets a
better battery as it runs at 60Hz v 120Hz on the Pros.', 'Good point']]
5
6 ['Love this phone 3 days of having and the battery last long! I had the 11pro for 3 years it's was all messed up and cracked so I had to
get another phone! I was going to get the pro max but they was out of stock I'm impatient and just wanted a fresh screen!', '\t3',
"\t['🥳🥳 <br>👉 THANKS FOR WATCHING YOU HAVE BEEN SELECTED AMONG MY LUCKY WINNERS MESSAGE ME ON TELEGRAM APP TO CLAIM YOUR PRIZE.....']]
7
8 ['This phone covers a niche audience where they want a vanilla 14 in a bigger form factor without the need for the Pro features. This is a
```