

#Problem Statement

"""

We have a huge number of comments from YouTube for a latest trailer from a worldwide production house, you as an AI service provider are supposed to analyse all the comments on that trailer, get the sentiment and the score, and give a consolidated report for that trailer about how it might perform on the box office.

"""

#Tools and high level steps that will be used in the project.

"""

1. Get your comments from the Youtube trailer. One option would manually getting the comments, get them directly from the API, you can also load it in your colab.

2. Setup your colab to do the job for you, you will need to install the required libraries.

PyTorch - torch

HuggingFace - transformers

NLTK - nltk

VADER - sentiment.vader

"""

"""

1. Do all the necessary imports

2. Create a function for removing stop words

3. Create a function to calculate the sentiment score and the sentiment (positive/negative)

4. Loop through the Comments that you will get from your input excel file

4.5 Segregate the words into positive and negative, so you can make a word cloud at the end

5. Calculate all the sentiments in loop and return only one final result

"""

!pip install torch

```

Requirement already satisfied: torch in /usr/local/lib/python3.11/dist-packages (2.6.0+cu124)
Requirement already satisfied: filelock in /usr/local/lib/python3.11/dist-packages (from torch) (3.18.0)
Requirement already satisfied: typing-extensions>=4.10.0 in /usr/local/lib/python3.11/dist-packages (from torch) (4.12.2)
Requirement already satisfied: networkx in /usr/local/lib/python3.11/dist-packages (from torch) (3.4.2)
Requirement already satisfied: Jinja2 in /usr/local/lib/python3.11/dist-packages (from torch) (3.1.6)
Requirement already satisfied: fsspec in /usr/local/lib/python3.11/dist-packages (from torch) (2025.3.0)
Collecting nvidia-cuda-nvrtc-cu12==12.4.127 (from torch)
  Downloading nvidia_cuda_nvrtc_cu12-12.4.127-py3-none-manylinux2014_x86_64.whl.metadata (1.5 kB)
Collecting nvidia-cuda-runtime-cu12==12.4.127 (from torch)
  Downloading nvidia_cuda_runtime_cu12-12.4.127-py3-none-manylinux2014_x86_64.whl.metadata (1.5 kB)
Collecting nvidia-cuda-cupti-cu12==12.4.127 (from torch)
  Downloading nvidia_cuda_cupti_cu12-12.4.127-py3-none-manylinux2014_x86_64.whl.metadata (1.6 kB)
Collecting nvidia-cudnn-cu12==9.1.0.70 (from torch)
  Downloading nvidia_cudnn_cu12-9.1.0.70-py3-none-manylinux2014_x86_64.whl.metadata (1.6 kB)
Collecting nvidia-cublas-cu12==12.4.5.8 (from torch)
  Downloading nvidia_cublas_cu12-12.4.5.8-py3-none-manylinux2014_x86_64.whl.metadata (1.5 kB)
Collecting nvidia-cufft-cu12==11.2.1.3 (from torch)
  Downloading nvidia_cufft_cu12-11.2.1.3-py3-none-manylinux2014_x86_64.whl.metadata (1.5 kB)
Collecting nvidia-curand-cu12==10.3.5.147 (from torch)
  Downloading nvidia_curand_cu12-10.3.5.147-py3-none-manylinux2014_x86_64.whl.metadata (1.5 kB)
Collecting nvidia-cusolver-cu12==11.6.1.9 (from torch)
  Downloading nvidia_cusolver_cu12-11.6.1.9-py3-none-manylinux2014_x86_64.whl.metadata (1.6 kB)
Collecting nvidia-cusparselt-cu12==0.6.2 (from torch)
  Downloading nvidia_cusparselt_cu12-0.6.2-py3-none-manylinux2014_x86_64.whl.metadata (1.6 kB)
Requirement already satisfied: nvidia-nccl-cu12==2.21.5 in /usr/local/lib/python3.11/dist-packages (from torch) (2.21.5)
Requirement already satisfied: nvidia-nvtx-cu12==12.4.127 in /usr/local/lib/python3.11/dist-packages (from torch) (12.4.127)
Collecting nvidia-nvjitlink-cu12==12.4.127 (from torch)
  Downloading nvidia_nvjitlink_cu12-12.4.127-py3-none-manylinux2014_x86_64.whl.metadata (1.5 kB)
Requirement already satisfied: triton==3.2.0 in /usr/local/lib/python3.11/dist-packages (from torch) (3.2.0)
Requirement already satisfied: sympy==1.13.1 in /usr/local/lib/python3.11/dist-packages (from torch) (1.13.1)
Requirement already satisfied: mpmath<1.4, >=1.1.0 in /usr/local/lib/python3.11/dist-packages (from sympy==1.13.1->torch) (1.3.0)
Requirement already satisfied: MarkupSafe>=2.0 in /usr/local/lib/python3.11/dist-packages (from Jinja2->torch) (3.0.2)
Downloading nvidia_cublas_cu12-12.4.5.8-py3-none-manylinux2014_x86_64.whl (363.4 MB)
 363.4/363.4 MB 4.9 MB/s eta 0:00:00
Downloading nvidia_cuda_cupti_cu12-12.4.127-py3-none-manylinux2014_x86_64.whl (13.8 MB)
 13.8/13.8 MB 90.9 MB/s eta 0:00:00
Downloading nvidia_cuda_nvrtc_cu12-12.4.127-py3-none-manylinux2014_x86_64.whl (24.6 MB)
 24.6/24.6 MB 68.2 MB/s eta 0:00:00
Downloading nvidia_cuda_runtime_cu12-12.4.127-py3-none-manylinux2014_x86_64.whl (883 kB)
 883.7/883.7 kB 41.1 MB/s eta 0:00:00
Downloading nvidia_cudnn_cu12-9.1.0.70-py3-none-manylinux2014_x86_64.whl (664.8 MB)
 664.8/664.8 MB 1.5 MB/s eta 0:00:00
Downloading nvidia_cufft_cu12-11.2.1.3-py3-none-manylinux2014_x86_64.whl (211.5 MB)
 211.5/211.5 MB 4.6 MB/s eta 0:00:00
Downloading nvidia_curand_cu12-10.3.5.147-py3-none-manylinux2014_x86_64.whl (56.3 MB)
 56.3/56.3 MB 12.8 MB/s eta 0:00:00
Downloading nvidia_cusolver_cu12-11.6.1.9-py3-none-manylinux2014_x86_64.whl (127.9 MB)
 127.9/127.9 MB 7.3 MB/s eta 0:00:00
Downloading nvidia_cusparselt_cu12-0.6.2-py3-none-manylinux2014_x86_64.whl (207.5 MB)
 207.5/207.5 MB 5.5 MB/s eta 0:00:00

```

Downloading nvidia_nvjitlink_cu12-12.4.127-py3-none-manylinux2014_x86_64.whl (21.1 MB)

21.1/21.1 MB 52.9 MB/s eta 0:00:00

Installing collected packages: nvidia-nvjitlink-cu12, nvidia-curand-cu12, nvidia-cufft-cu12, nvidia-cuda-runtime-cu12, nvidia-cuda-

Attempting uninstall: nvidia-nvjitlink-cu12

Found existing installation: nvidia-nvjitlink-cu12 12.5.82

Uninstalling nvidia-nvjitlink-cu12-12.5.82:

!pip install transformers

```

Requirement already satisfied: transformers in /usr/local/lib/python3.11/dist-packages (4.50.0)
Requirement already satisfied: filelock in /usr/local/lib/python3.11/dist-packages (from transformers) (3.18.0)
Requirement already satisfied: huggingface-hub<1.0,>=0.26.0 in /usr/local/lib/python3.11/dist-packages (from transformers) (0.29.3)
Requirement already satisfied: numpy>=1.17 in /usr/local/lib/python3.11/dist-packages (from transformers) (2.0.2)
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.11/dist-packages (from transformers) (24.2)
Requirement already satisfied: pyyaml>=5.1 in /usr/local/lib/python3.11/dist-packages (from transformers) (6.0.2)
Requirement already satisfied: regex!=2019.12.17 in /usr/local/lib/python3.11/dist-packages (from transformers) (2024.11.6)
Requirement already satisfied: requests in /usr/local/lib/python3.11/dist-packages (from transformers) (2.32.3)
Requirement already satisfied: tokenizers<0.22,>=0.21 in /usr/local/lib/python3.11/dist-packages (from transformers) (0.21.1)
Requirement already satisfied: safetensors>=0.4.3 in /usr/local/lib/python3.11/dist-packages (from transformers) (0.5.3)
Requirement already satisfied: tqdm>=4.27 in /usr/local/lib/python3.11/dist-packages (from transformers) (4.67.1)
Requirement already satisfied: fsspec>=2023.5.0 in /usr/local/lib/python3.11/dist-packages (from huggingface-hub<1.0,>=0.26.0->transformers) (2024.11.6)
Requirement already satisfied: typing-extensions>=3.7.4.3 in /usr/local/lib/python3.11/dist-packages (from huggingface-hub<1.0,>=0.26.0->transformers) (4.12.0)
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.11/dist-packages (from requests->transformers) (3.3.0)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.11/dist-packages (from requests->transformers) (3.10)
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.11/dist-packages (from requests->transformers) (2.3.0)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.11/dist-packages (from requests->transformers) (2025.1.1)

```

!pip install nltk

```

Requirement already satisfied: nltk in /usr/local/lib/python3.11/dist-packages (3.9.1)
Requirement already satisfied: click in /usr/local/lib/python3.11/dist-packages (from nltk) (8.1.8)
Requirement already satisfied: joblib in /usr/local/lib/python3.11/dist-packages (from nltk) (1.4.2)
Requirement already satisfied: regex>=2021.8.3 in /usr/local/lib/python3.11/dist-packages (from nltk) (2024.11.6)
Requirement already satisfied: tqdm in /usr/local/lib/python3.11/dist-packages (from nltk) (4.67.1)

```

!pip install vaderSentiment

```

Collecting vaderSentiment
  Downloading vaderSentiment-3.3.2-py2.py3-none-any.whl.metadata (572 bytes)
Requirement already satisfied: requests in /usr/local/lib/python3.11/dist-packages (from vaderSentiment) (2.32.3)
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.11/dist-packages (from requests->vaderSentiment) (3.3.0)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.11/dist-packages (from requests->vaderSentiment) (3.10)
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.11/dist-packages (from requests->vaderSentiment) (2.3.0)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.11/dist-packages (from requests->vaderSentiment) (2025.1.1)
  Downloading vaderSentiment-3.3.2-py2.py3-none-any.whl (125 kB)
    126.0/126.0 kB 2.5 MB/s eta 0:00:00
Installing collected packages: vaderSentiment
Successfully installed vaderSentiment-3.3.2

```

```

import nltk
#from nltk.sentiment import SentimentIntensityAnalyzer
nltk.download('vader_lexicon')
nltk.download('stopwords')
nltk.download('punkt_tab')

```

```

[nltk_data] Downloading package vader_lexicon to /root/nltk_data...
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data] Unzipping corpora/stopwords.zip.
[nltk_data] Downloading package punkt_tab to /root/nltk_data...
[nltk_data] Unzipping tokenizers/punkt_tab.zip.
True

```

```
from transformers import pipeline
```

```

from vaderSentiment.vaderSentiment import SentimentIntensityAnalyzer
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize

```

```

sia = SentimentIntensityAnalyzer()
stop_words = stop_words = set(stopwords.words('english'))
classifier = pipeline("sentiment-analysis", model="distilbert/distilbert-base-uncased-finetuned-sst-2-english")

```

```

/usr/local/lib/python3.11/dist-packages/huggingface_hub/utils/_auth.py:94: UserWarning:
The secret `HF_TOKEN` does not exist in your Colab secrets.
To authenticate with the Hugging Face Hub, create a token in your settings tab (https://huggingface.co/settings/tokens), set it as :
You will be able to reuse this secret in all of your notebooks.
Please note that authentication is recommended but still optional to access public models or datasets.
warnings.warn(
config.json: 100% 629/629 [00:00<00:00, 36.4kB/s]

model.safetensors: 100% 268M/268M [00:01<00:00, 138MB/s]

tokenizer_config.json: 100% 48.0/48.0 [00:00<00:00, 3.21kB/s]

vocab.txt: 100% 232k/232k [00:00<00:00, 3.29MB/s]

Device set to use cpu

import pandas as pd
df = pd.read_excel("/content/SnowWhite Comments.xlsx")

comments = []
comments = df['Comments'].tolist()

def remove_stopwords(raw_comment):
    tokenized_comment = word_tokenize(raw_comment)
    processed_comment = [ word for word in tokenized_comment if word.lower() not in stop_words]
    return ' '.join(processed_comment)

def get_comment_sentiment_details(raw_comment):
    processed_comment = remove_stopwords(raw_comment)

    words = processed_comment.split()
    positive_words = ""
    negative_words = ""
    comment_sentiment = "" #Either POSITIVE or NEGATIVE

    sentence_score_temp = sia.polarity_scores(processed_comment)

    abs_sentence_score = abs(sentence_score_temp['compound']) #absolute means if I have -3.4 -> 3.4
    sentiment_label = classifier(processed_comment)
    comment_sentiment = sentiment_label[0]['label']

    if abs_sentence_score == 0:
        comment_sentiment = "NEUTRAL"

    if comment_sentiment == "NEGATIVE":
        sentence_score = abs_sentence_score * -1
        for word in words:
            word_sentiment = sia.polarity_scores(word)
            if word_sentiment['compound'] < 0:
                negative_words += word + " "

    elif comment_sentiment == "POSITIVE":
        sentence_score = abs_sentence_score
        for word in words:
            word_sentiment = sia.polarity_scores(word)
            if word_sentiment['compound'] > 0:
                positive_words += word + " "

    else:
        sentence_score = abs_sentence_score

    return positive_words, negative_words, sentence_score, comment_sentiment

positive_words = ""
negative_words = ""
neu_count = 0

pos_values_list = []
neg_values_list = []
avg_pos_score = 0
avg_neg_score = 0

for comment in comments:
    pw, nw, ss, cs = get_comment_sentiment_details(comment)
    positive_words += pw+ " " #storing pw from each comment into our central positive words
    negative_words += nw+ " "

    if cs == "NEGATIVE":
        neg_values_list.append(ss)

```

```

elif cs == "POSITIVE":
    pos_values_list.append(ss)
else:
    neu_count+=1

try:
    avg_pos_score = sum(pos_values_list) / len(pos_values_list)
    avg_neg_score = sum(neg_values_list) / len(neg_values_list)
except ZeroDivisionError:
    if len(pos_values_list) == 0 or len(neg_values_list) == 0:
        avg_pos_score = 0
        avg_neg_score = 0

final_score = (avg_pos_score + avg_neg_score) / (len(pos_values_list) + len(neg_values_list))

print(final_score)

0.0023106914414414447

```

positive_words

```

' Yay hope beauty true beauty beautiful yes beautiful Pretty hope credit brave open 🥰 greatest
true prettier 🥰 likes great ❤️ Admit lover hope hope win supportive masterpiece romantic Marv
admit advice heavenly save great want original original Hell giving love thank Take great

```

negative_words

```

'poison broke hate mess 🤬 insult stop stop disappointed dislike dislikes bomb destroy cruelty bad bad
seriously madness WORST DOOM DROP DISLIKE LEAVE stop propaganda CGI... 🤡 mad 🤡 🤡 🤡 worst Bad Bad poor
chuck empty poor Nah scream cut 🤡 🤡 lost lowest destroy dumpster fire 🤡 🤡 🤡 🤡 🤡 🤡 🤡 🤡 🤡 🤡

```

avg_pos_score

```
0.5963791666666668
```

avg_neg_score

```
-0.42538799999999999
```

!pip install wordcloud

```

Requirement already satisfied: wordcloud in /usr/local/lib/python3.11/dist-packages (1.9.4)
Requirement already satisfied: numpy>=1.6.1 in /usr/local/lib/python3.11/dist-packages (from wordcloud) (2.0.2)
Requirement already satisfied: pillow in /usr/local/lib/python3.11/dist-packages (from wordcloud) (11.1.0)
Requirement already satisfied: matplotlib in /usr/local/lib/python3.11/dist-packages (from wordcloud) (3.10.0)
Requirement already satisfied: contourpy>=1.0.1 in /usr/local/lib/python3.11/dist-packages (from matplotlib->wordcloud) (1.3.1)
Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.11/dist-packages (from matplotlib->wordcloud) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in /usr/local/lib/python3.11/dist-packages (from matplotlib->wordcloud) (4.56.0)
Requirement already satisfied: kiwisolver>=1.3.1 in /usr/local/lib/python3.11/dist-packages (from matplotlib->wordcloud) (1.4.8)
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.11/dist-packages (from matplotlib->wordcloud) (24.2)
Requirement already satisfied: pyparsing>=2.3.1 in /usr/local/lib/python3.11/dist-packages (from matplotlib->wordcloud) (3.2.1)
Requirement already satisfied: python-dateutil>=2.7 in /usr/local/lib/python3.11/dist-packages (from matplotlib->wordcloud) (2.8.2)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.11/dist-packages (from python-dateutil>=2.7->matplotlib->wordcloud) (1.16.0)

```

```

from wordcloud import WordCloud
import matplotlib.pyplot as plt

```

```


print("positives")
wordcloud_positive = WordCloud(width=800, height=400, background_color='white').generate(positive_words)

```

```


plt.figure(figsize=(10,5))
plt.imshow(wordcloud_positive, interpolation='bilinear')
plt.axis('off')
plt.show()

```

 positives

```
print("negatives")
wordcloud_negative = WordCloud(width=800, height=400, background_color='white').generate(negative_words)

plt.figure(figsize=(10,5))
plt.imshow(wordcloud_negative, interpolation='bilinear')
plt.axis('off')
plt.show()
```

 negatives