

WARM UP PROJECT MODIFIED FILE

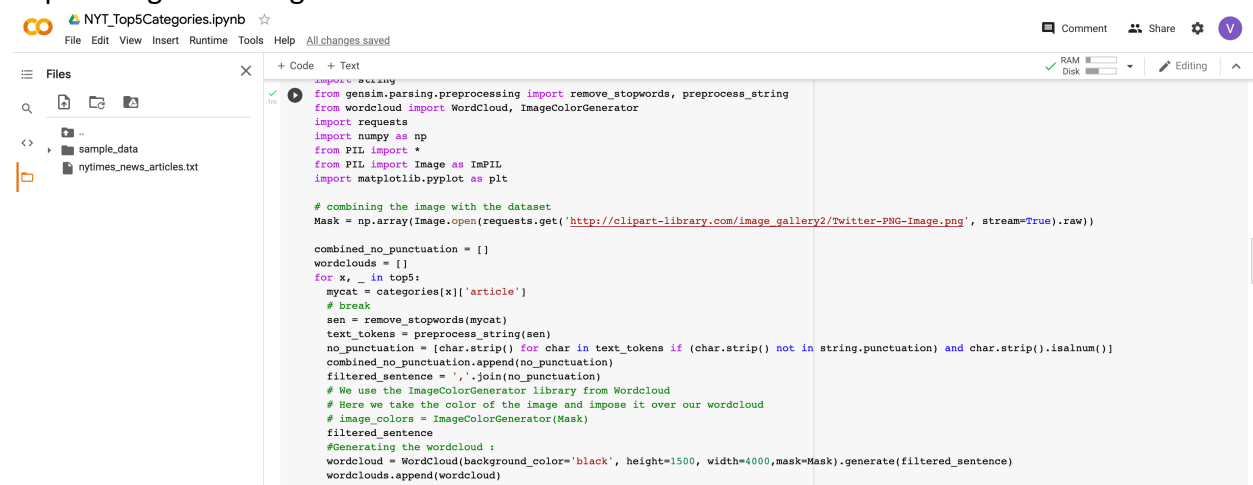
FOR TASK 2

It states that we have to generate word cloud for Top 5 Categories. Since it did not state combined wordcloud, I did it individually. As discussed with TA, we would like to modify our code and we are sorry for misunderstanding. Next time we will make sure that we understand the question thoroughly to get desired results.

I have not referred any others code or asked TA to show us the solution. I have worked on my own to generate combined word cloud of Top 5 Categories

Here is the code for generating combined wordcloud.

Finding individual wordclouds, as in warm up project uploaded file, I have clearly stated each Top 5 categories and generated the word cloud.



```
NYT_Top5Categories.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment Share Settings V
RAM Disk Editing ^
Files
sample_data
nytimes_news_articles.txt
+ Code + Text
from gensim.parsing.preprocessing import remove_stopwords, preprocess_string
from wordcloud import WordCloud, ImageColorGenerator
import requests
import numpy as np
from PIL import *
from PIL import Image as ImPIL
import matplotlib.pyplot as plt

# combining the image with the dataset
Mask = np.array(Image.open(requests.get('http://clipart-library.com/image_gallery2/Twitter-PNG-Image.png', stream=True).raw))

combined_no_punctuation = []
wordclouds = []
for x, _ in top5:
    mycat = categories[x]['article']
    # break
    sen = remove_stopwords(mycat)
    text_tokens = preprocess_string(sen)
    no_punctuation = [char.strip() for char in text_tokens if (char.strip() not in string.punctuation) and char.strip().isalnum()]
    combined_no_punctuation.append(no_punctuation)
    filtered_sentence = ','.join(no_punctuation)
    # We use the ImageColorGenerator library from Wordcloud
    # Here we take the color of the image and impose it over our wordcloud
    # image_colors = ImageColorGenerator(Mask)
    filtered_sentence
    #Generating the wordcloud
    wordcloud = WordCloud(background_color='black', height=1500, width=4000,mask=Mask).generate(filtered_sentence)
    wordclouds.append(wordcloud)
```

Combining Top 5 Categories Wordcloud



```
NYT_Top5Categories.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment Share Settings V
RAM Disk Editing ^
Files
sample_data
nytimes_news_articles.txt
+ Code + Text
[8] final_punctuation = []
for no_punctuation in combined_no_punctuation:
    for s_char in no_punctuation:
        final_punctuation.append(s_char)
final_sentence = ','.join(final_punctuation)
wordcloud = WordCloud(background_color='black', height=1500, width=4000,mask=Mask).generate(final_sentence)
```

Output of Wordcloud Combining TOP 5 Categories

[illegible]