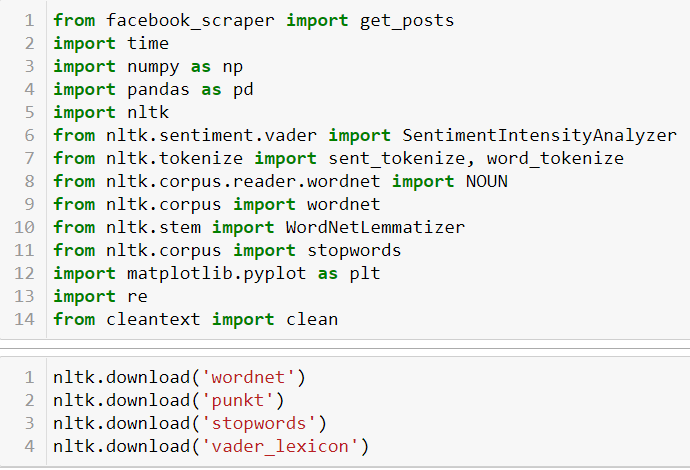
**Social Web Analytics Project**

**Facebook:**

In today’s age Facebook is a huge social media platform, where millions of people interact not just with each but also with popular brands. Coca-Cola, Pepsi, Sprite are 3 such companies that are active on Facebook and goal is to find out how much their activity is liked by the users based on sentiment analysis of the scrapped data from their respective verified pages.

**Importing Libraries:**

Firstly, importing the required libraries. The library of main focus is facebook\_scraper that can easily scrape any amount of post data from any Facebook page/group. For sentiment analysis Vader library is imported which gives positive, negative, neutral, and compound sentiment for each text word/sentence.



**Data Scraping:**

get\_posts function from facebook\_scraper library fetches post data from the given facebook page IDs: “cocacolaindia”, “PepsiIndia”, “spritein”.

For the get\_posts function: -

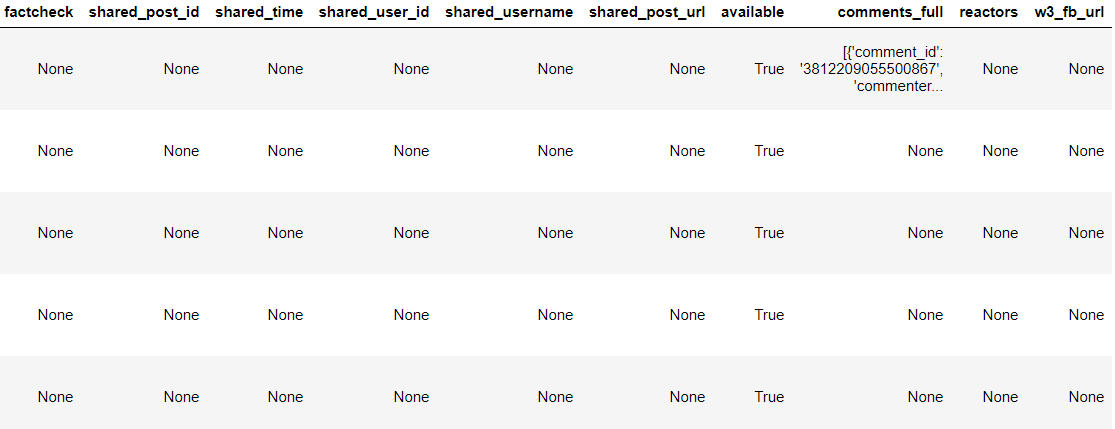
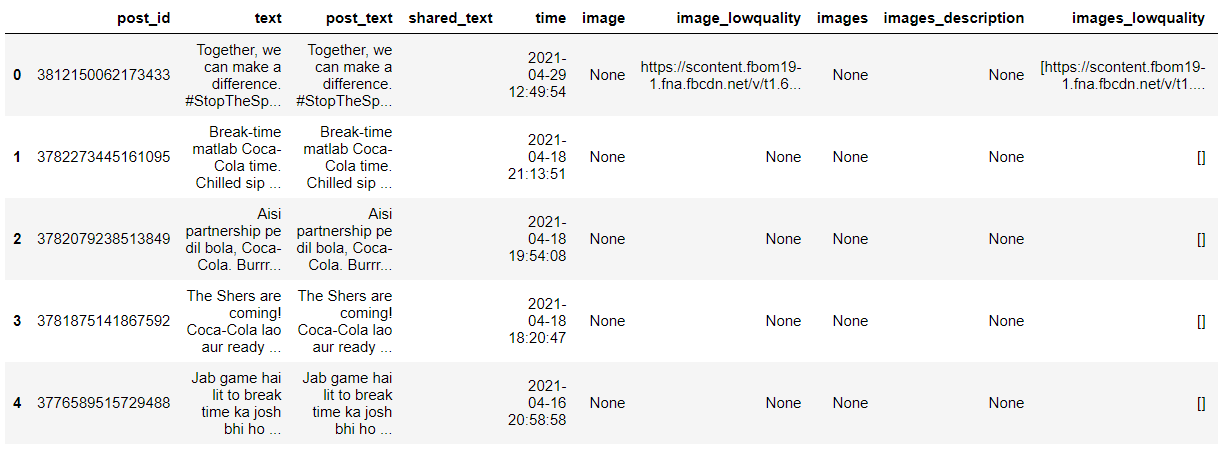
* **group**: group id, to scrape groups instead of pages. Default is None.
* **pages**: how many pages of posts to request, the first 2 pages may have no results, so try with a number greater than 2. Default is 10.
* **timeout**: how many seconds to wait before timing out. Default is 5.
* **credentials**: tuple of user and password to login before requesting the posts. Default is None.
* **extra\_info**: bool, if true the function will try to do an extra request to get the post reactions. Default is False.
* **youtube\_dl**: bool, use Youtube-DL for (high-quality) video extraction. You need to have youtube-dl installed on your environment. Default is False.
* **post\_urls**: list, URLs or post IDs to extract posts from. Alternative to fetching based on username.
* **cookies**: One of:
  + The path to a file containing cookies in Netscape or JSON format
  + A CookieJar
  + A dictionary that can be converted to a CookieJar with cookiejar\_from\_dict
* **options**: Dictionary of options. Set options={"comments": True} to extract comments, set options={"reactors": True} to extract the people reacting to the post. Both comments and reactors can also be set to a number to set a limit for the number of comments/reactors to retrieve. The default limit for comments is 5000 and the default limit for reactors is 3000. Set options={"progress": True} to get a tqdm progress bar while extracting comments and replies. Set options={"allow\_extra\_requests": False} to disable making extra requests when extracting post data (required for some things like full text and image links). Set options={"posts\_per\_page": 200} to request 200 posts per page. The default is 4.



**Dataset:**

After scraping the data for Coca-Cola, Pepsi and Sprite we convert the scraped data into Data Frames for ease of use when performing analysis. Given, below are the dataset obtained from scraping 202 posts of 3 brands that include likes, comments, date of post and the other metrics.

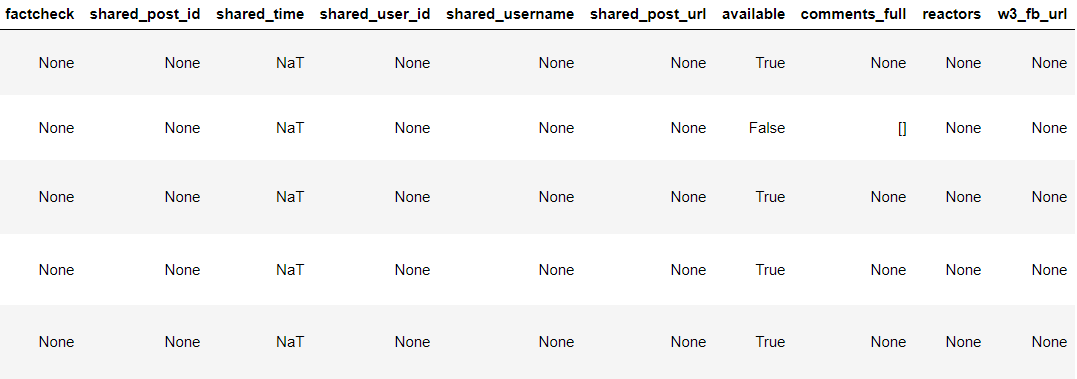
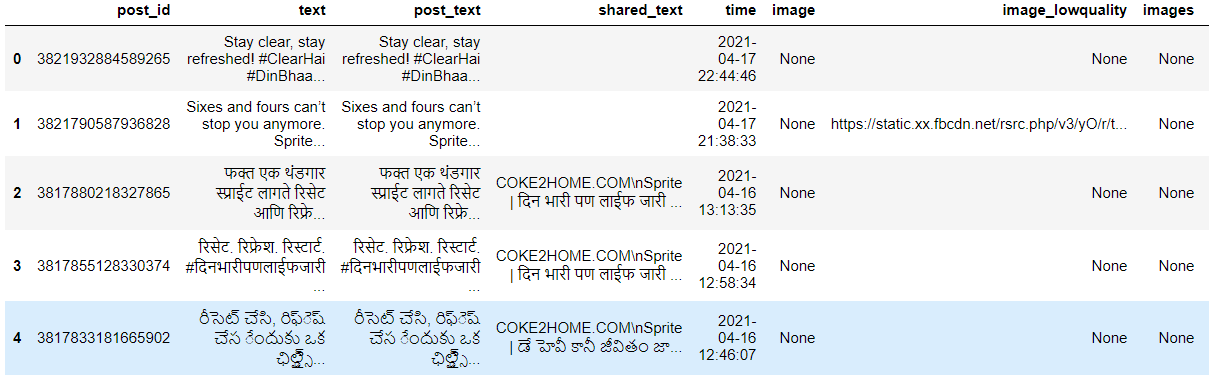
**Coca-Cola:**



**Pepsi:**



**Sprite:**



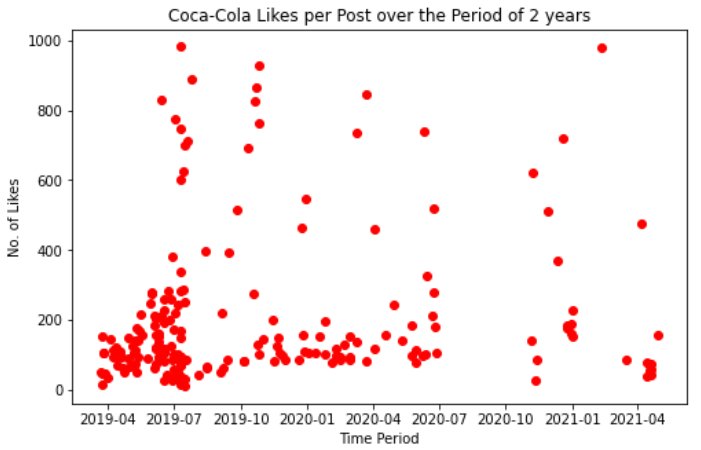
**Analysis:**

Coming to the analysis part, before we analyse the sentiments from the scraped data we need to understand how much the 3 brands are “liked” by the people and how much are interactive on Facebook platform. For this purpose, 2 different plots are to be made for each brand.

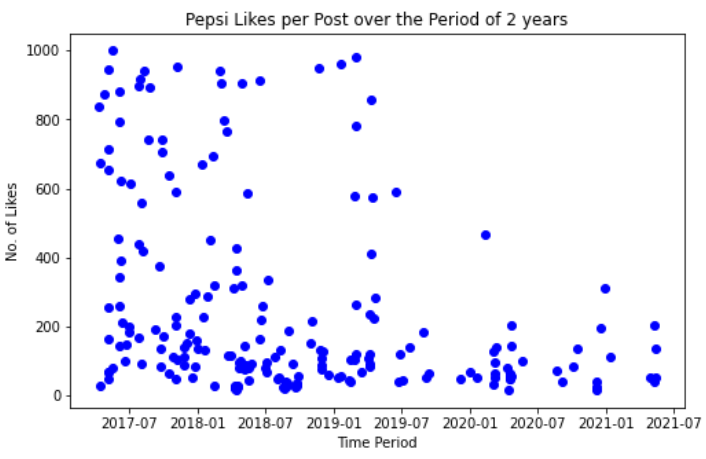
**Part-1**

Firstly, a graph that plots the number of likes on each post for 202 posts over the period. (The period for each brand may differ since none have any fixed time period for posting a post.)

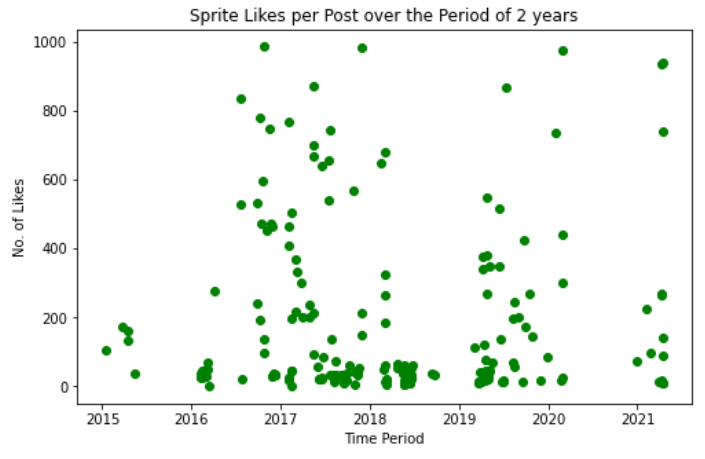
**Coca-Cola:**



**Pepsi:**



**Sprite:**



As we can see from the above 3 graphs:

1. Coca-Cola has had good number of likes on each post since 2019 but they seem to have dropped recently, however the random spikes in number of likes are still present which might be due to the launch of any of their unique advertisements that attracted the people.
2. Pepsi on the other hand has captured people mind and has been dominating in the number of likes since 2017 but, recently it has faced sharp decline with no new spikes.
3. Sprite which hasn’t been much liked by people since 2015 with spikes in 2017 and 2019 has been liked bit more compared to previous posts.

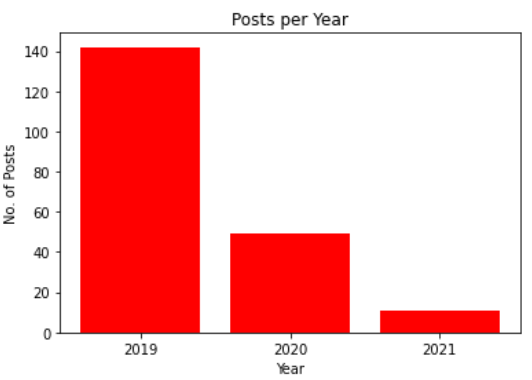
**Part-2**

Now, for the second part we make a plot for number of posts each year for the 3 brands to measure their consumer interaction.

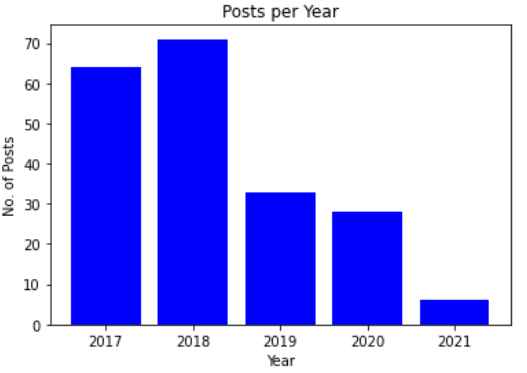
As, we can see below in the 3 graphs:

1. Number of posts which were extremely high in 2019 (140+ posts) has come down to 40-60 posts in 2020 and even lower in 2021. This shows Coca-Cola has suffered a huge blow in customer engagement on social media which might be due to the covid situation.
2. On a similar basis Pepsi’s number of posts has halved once in 2019 and again in 2021 till date. This shows even Pepsi has lowered its customer engagement on Facebook platform.
3. Sprite has always had lowest number of posts compared to Pepsi and Coca-Cola even during its peak period (2016-2019). However, recently in 2021 till date it has most number of posts compared to Pepsi and Coca-Cola.

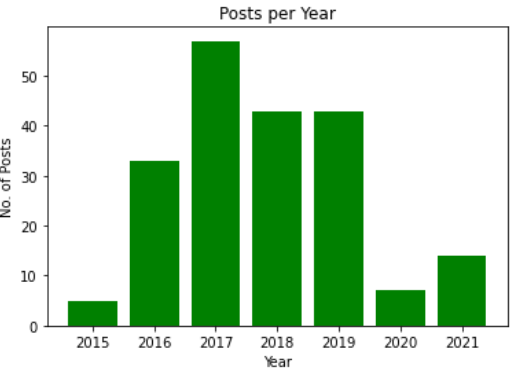
**Coca-Cola:**



**Pepsi:**



**Sprite:**



**Part-3**

Finally, coming to the most important part that is sentiment analysis of the comments posted on posts of each of these three brands.

But, firstly for that we extract only the comments portion from the posts data which was scrapped earlier for all the three brands. Once, that is done we clean the “comments\_text” using a “clean\_text” function created specially to clean the comments by:

1. Changing case of all words in the comment to lower case.
2. Using cleantext library to clean the comment.
3. Using regular expression to remove the unwanted letters, numbers, special characters, emojis etc.
4. Tokenizing the comments
5. Using stopwords from nltk library to remove the frequent occurrences of common articles.
6. Lastly, lemmatizing and joining the words.

After applying the above cleaning to the comment’s column, we obtain a cleaned\_comments column which is a filtered and clean text ready for sentiment analysis.

VADER uses a combination of A sentiment lexicon which is a list of lexical features (e.g., words) which are generally labeled according to their semantic orientation as either positive or negative.

Now, we use the Vader library’s Sentiment Intensity Analyzer on the cleaned\_comments which gives the intensity of the sentiment behind the comments that is Positive, Negative, Neutral and Compound. Hence, after storing these values we obtain new columns as shown below in the comments table marking the sentiment intensity for each comment.

As, it is visible in the data below Positive, Negative, Neutral are percentage values lying between 0 and 1 that shows the positivity, negativity, and neutrality of each comment.

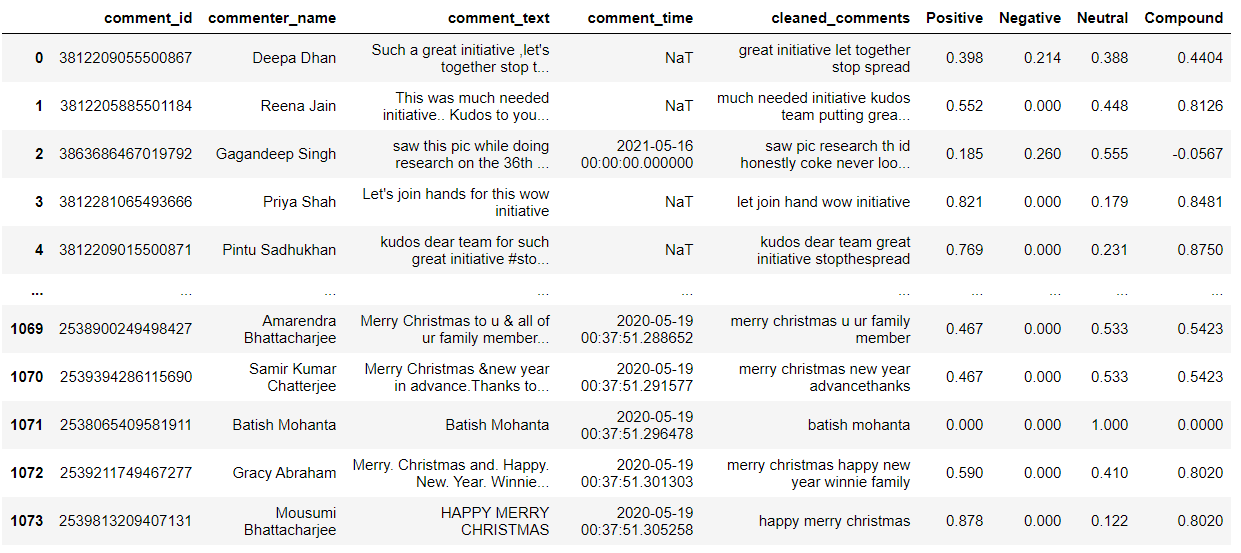
Thus, the sum of Positive, Negative and Neutral is always equal to 1.

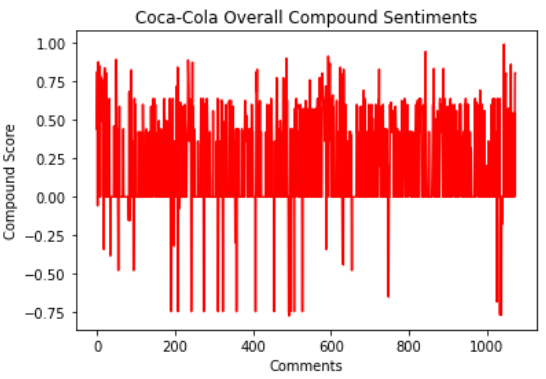
The Compound score is a metric that calculates the sum of all the lexicon ratings which have been normalized between -1 (extreme negative) and +1 (extreme positive)**.**

Now, we plot the Compound Sentiments for each of the comments for all three brands to get an insight of how positive or negative people’s sentiments are towards the brand.

**Coca-Cola:**

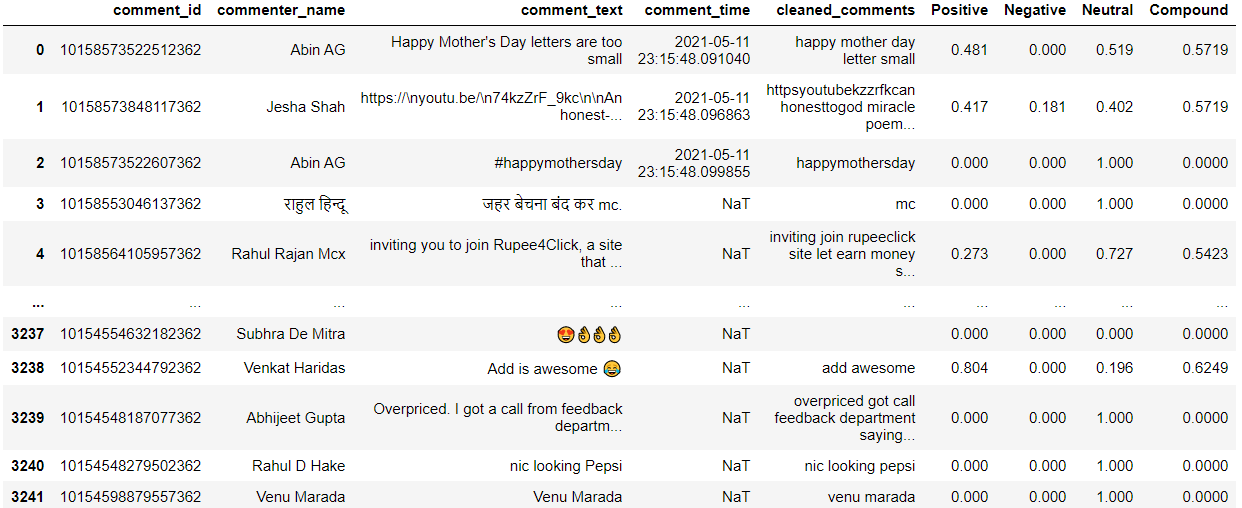
As, you can see in the below plot Coca-Cola has 1074 comments over 202 posts with mostly positive sentiments which highlights the positive image of Coca-Cola in front of people on social media.

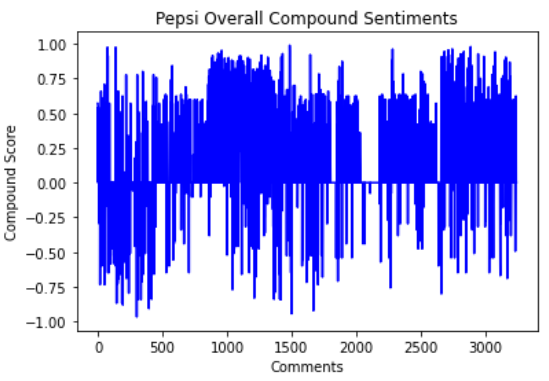




**Pepsi:**

In the below plot of Pepsi even though it has had humongous response to its posts with 3242 comments over 202 posts it is still struggling to obtain positive sentiments as it has a large mix of both positive and negative sentiments. However, it is successful in gaining consumer engagement seeing the huge response to its posts which is a big plus due to which they have better chances of improving their products as per feedbacks received.

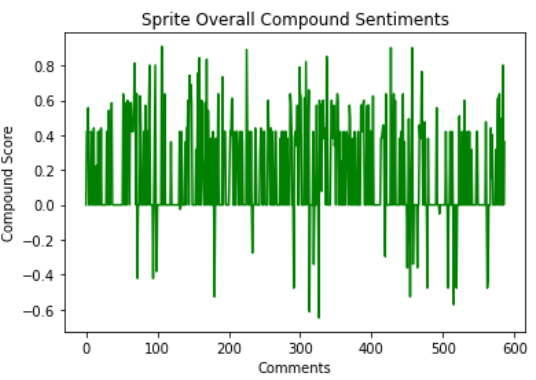




**Sprite:**

As seen earlier, compared to Coca-Cola and Pepsi, Sprite has very low customer engagement with only 587 comments over 202 posts. Though on the plus side majority of those comments are positive comments which means even with low consumer engagement Sprite is still liked by people.





**Result:**

Overall with a few trade offs here and there Pepsi is the best brand on Facebook platform as it is able engage with people on a much larger scale compared to Coca-Cola and Sprite even if there is a positive – negative sentiment mix in response from people, a fast selling brand like Pepsi is better as it interacts more with its consumers.