

## **CIS7029 Social Media Analytics for Business**

**Module Leader: Dr Esyin Chew**

**Title: Coca Cola vs Pepsi  
Social Media Analytics**

By

Rohit Saini  
MSc Data Science (Internship)  
St20183550

## **Biography**

As a Data Analyst of an upcoming beverage brand my role in this project is to analyse the data of two different brands within the beverage market. As an analyst I'm going to compare top two competitors of beverage industry i.e., Coca Cola and Pepsi. In this project I'm going to retrieve and gather social media data of both the competitor using different tools and techniques, organize it in a meaningful charts and graphs and then use these graphs and charts to find some meaningful conclusion which can help the upcoming business.

## **Abstract**

### **Purpose**

The main purpose of the research is to find who is the winner on social media between two giant beverage brands competitor i.e., Coca Cola and Pepsi. Then take some ideas from the winner for our own business purposes.

### **Methodology**

Research is done with the help of many methods and tools which are used for extracting social media data i.e., likes, comments, shares and retweets from both Coca Cola and Pepsi official page. Python coding is done for extracting finance information of both brands from Wikipedia using few libraries. Data sources like Excel and visualisation tool like Tableau is used for analysing the extracted data. Some google chrome extensions are also used. A method for analysing sentiments of twitter tweets is done by using Jupyter Notebook and Twitter API.

### **Findings**

After analysing different types of data of both the brands it is concluded that overall Coca Cola has more social media engagement than Pepsi on almost all social media platforms. As social media has a capability to increase the success of the company, the giant Coca Cola beats Pepsi in terms of net income and net revenue in the case of beverages.

### **Business Implication**

The research is implemented in the business where social media engagement matters. The upcoming brand can take the ideas from the winning brand of this research (i.e., Coca Cola) to increase their value.

	Page No.
1. Biography	i
2. Abstract	i
3. Introduction	2
4. Design and Discussion	3
5. Visualisation and Results	7
6. Limitations	15
7. Conclusion and Recommendation	15
8. References	16
9. Appendix	18

## **Introduction**

Social media is a very big platform where communication is done between two or peoples, groups or organizations. In the terms of personal life, social media helps an individual to keep in touch with their friends and family. Social media has a great ability to grow and affect the success of the company or brand. For organizations or companies, it is helpful by sharing information of their business or their thoughts about their products with their customers. Social media allows people from different cities or countries of the world to connect with each other. An individual can update any information at any time using social media (Plume, Slade and Dwivedi, 2017).

Social media data is generally the information that is collected from social sites that can show how users view and engage with their profile. Social Data is very beneficial to marketers by looking up their customer reviews that can increase their sales. There are different types of social data which includes posts on Facebook, videos on YouTube, tweets on Twitter, photos and videos on Instagram. For Business purpose, Facebook and Twitter are the top most social platform used by the organization (Frankenfield, 2019).

In today's world brands are fighting out not only on stores but also in social media. In this research paper we are going to examine the strategy and engagement of two major brand competitors i.e., Coca Cola and Pepsi on social media. For a business plan, I am looking at both the brands to see from which we can takes ideas. The main prospective of this study is to discover which brand is more socially active and is the socially activeness helps in their business or not and then take ideas from the winning brand for our upcoming new to market beverage brand.

Coca cola is soft drink which is produced by The Coca Cola Company. It was introduced 134 years ago in United States. This iconic American Brand is recognized instantly everywhere in the world and sold in more than 200 countries (Wikipedia Contributors, 2019a).

Pepsi is also a soft drink which is produced by PepsiCo. It was introduced 127 years ago as “Brad’s Drink” but after 5 years it changes its name to “Pepsi Cola” and then after 63 years ago its name changed to “Pepsi”. Pepsi is also an American brand and is sold in many countries (Wikipedia Contributors, 2019b).

## **Design and Discussion**

Social media data is the biggest and most active data in today's generation. Social media data is the new way to understand people's human behavior. In this advance generation where technology is the key for almost every type of work the scientists and programmers are finding and evaluating different methods of automatically collecting, combining and analyzing the data (Batinica and Treleaven, 2014). A variety of tools, methods and API are used in this paper for data harvesting and analyzing. Let's have a brief description of tools used:

### **Octoparse**

Octoparse is a free data scrapping software used for scrapping different types of data without any coding. It is very user friendly and straight forward software. It turns the unstructured data from any webpage to a structured dataset (Team, 2020). For the purpose of this paper, this software is used for scrapping Twitter, Facebook and YouTube data. (See Appendix A1)

### **Data Miner and Scraper**

Data Miner and Scraper is a google chrome extension that helps the individuals to extract data from a website and save it to Excel spreadsheets. It basically transforms the HTML unstructured data into a structured tabular form (Data Miner, n.d.) (chrome.google.com, n.d.). In this paper both Data Miner and Scraper is used for extracting the last year finance information of both Coca Cola Co and Pepsi Co from Wikipedia. (See Appendix A2 and A3)

### **Python - Spyder IDE**

Python is an outstanding and efficient programming language used by many developers and scientist all over the world. Python can communicate with many servers including the live streaming servers. Data scientists use python for many purposes like storing the fetched data into a database, for analyzing or for making predictions (How To Developers: Using Python for the Cloud and Big Data Analytics, 2015). Spyder is an open-source fully integrated development environment used for programming in Python language (Educative: Interactive Courses for Software Developers, n.d.). It has multiple windows for the script and console for interactive use. It also has a help window with search by subject. It is so easy to write and run script in Spyder and has a logical way to create projects and manage them. For the purpose of this paper, Spyder is used for extracting year wise financial data of Coca Cola Co and PepsiCo and then save the extracted data to an Excel spreadsheet.

Main libraries used in python code while extracting data are:

1. Urllib.request – Urllib.request is python library used for open a particular webpage. The urllib.request contains a lot of functions but for the purpose of this research I've used the “urllib.request.urlopen” (docs.python.org, n.d.) .
2. Beautiful Soup – Beautiful soup is a python API used for data scraping. This API can scrap data from both XML and HTML files (Idris, 2016). It works with a parser to search, navigate or modify a parse tree (Richardson, 2019).

In this research paper, URLLIB.REQUEST and BEAUTIFUL SOUP works together for extracting data from the Coca Cola Co and PepsiCo Wikipedia page. Urllib.request.urlopen helps in opening the required page and then beautiful soup did their work by navigating and searching for the required data. (See Appendix A4)

## **Microsoft Excel**

Excel is the most common spreadsheets used for almost every big to small businesses. For a Data Analyst it is very useful to do some basic calculations. As a data analyst we can sort, filter and format the data. Charts and tables can also be made through Excel (Excel-easy.com, 2010). For the purpose of this research, I've used excel as a data source to save the scraped data for cleaning and then used the clean data for visualization. (See Appendix A5)

## **Tableau**

Tableau is a data visualization tool that enables us to create beautiful visualizations in the form of particular sheets, dashboards and stories to make better business development. Even non-technical users can also easily create customized dashboards and stories that provide vision to a wide range of information (Tableau, n.d.). For the purpose of this research paper, I've used tableau for visualizing the data into suitable charts and graphs. For twitter I've used packed bubble, for YouTube I've used treemaps, for comparing both companies' social media followers I've used side by side bars. (See Appendix A6 and Appendix A8)

## **Jupyter Notebook**

Jupyter notebook is a web application which is an open source that allows the individuals to create code, analyze data, visualize data, machine learning and much more. This notebook connects to many kernels at the same period to allow programming in different languages (By Jupyter It All Make Sense, 2018). In this research, this notebook is used for analyzing the twitter sentiments of a certain period of time with the help of twitter API and then draft it into graphical form for analyzing. Some libraries, packages and modules used for analyzing these sentiments are: (See Appendix A7)

1. Pandas – Pandas is a popular python-based data analysis toolkit. It can be used for loading and saving of data, column insertion and deletion, data sorting, data visualization and many more (Mckinney, 2018).
2. Matplotlib – Matplotlib is a python library used for plotting attractive 2-dimensional graphs. It consists of several plots like line, scatter, bar, histogram etc. It makes heavy use of NumPy and other extensions of code to provide a better performance even for large arrays (Mckinney, 2018).
3. Seaborn – Seaborn is another python visualization library that is based on Matplotlib. It provides a high level of attractiveness for statistical graphs (Mckinney, 2018).
4. Tweepy – Tweepy is a python package used for accessing Twitter API. It basically used for connecting python with twitter (Garcia, n.d.).
5. NLTK – There are few Natural Language Processing libraries in python but NLTK (Natural Language Tool Kit) is certainly the most well-known for NLP. NLTK comes with many lexical resources, utilities and bundled with many parsers (Perkins, 2010).

6. Textblob – Textblob is also a library in python for Natural language Processing like NLTK. But TextBlob is designed to gain more benefits than NLTK (Shubham Jain, 2018).

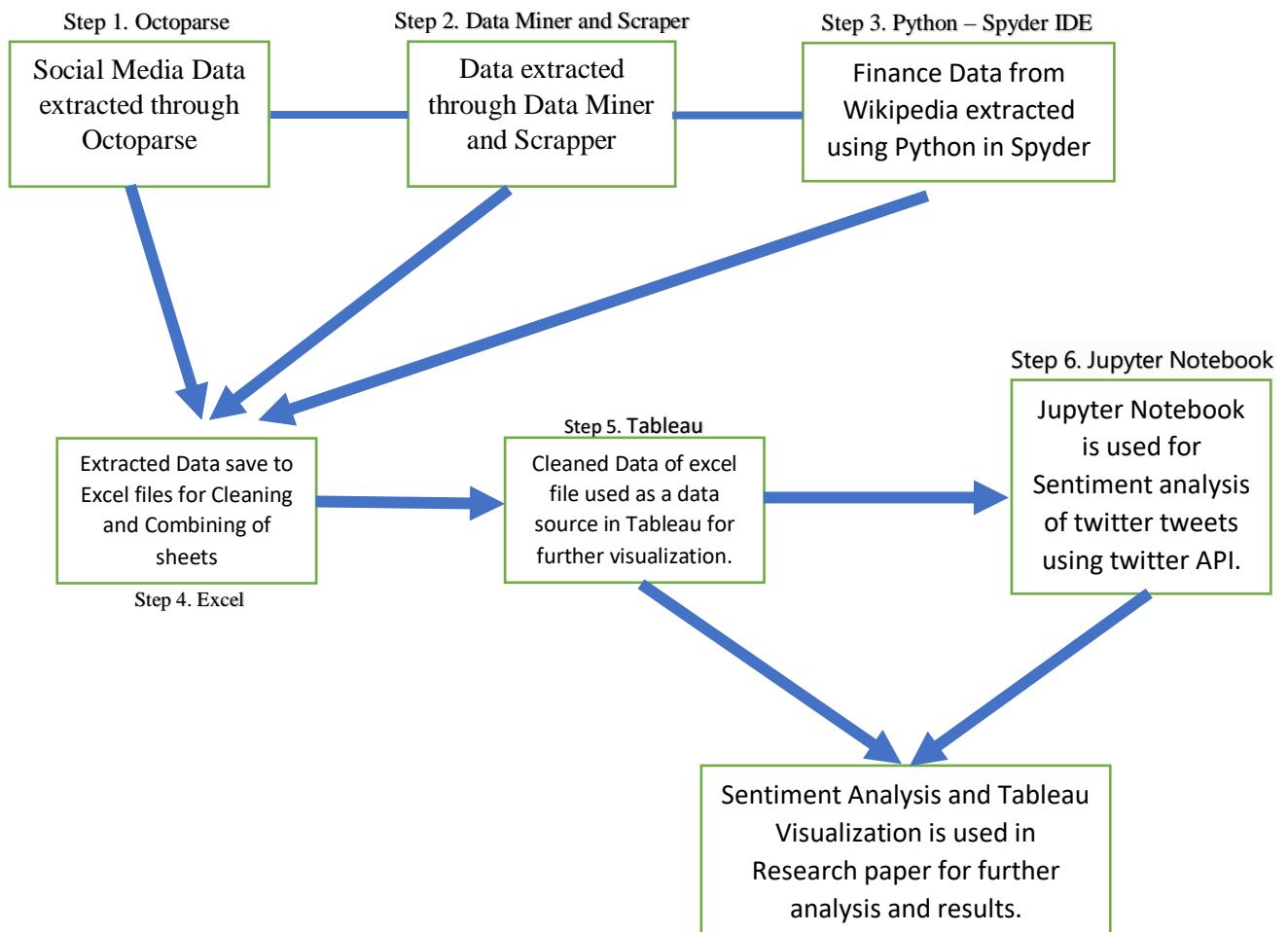
## **Online Tools**

Apart from Textblob and NLTK another online tool is used for sentiment analysis of twitter tweets i.e.; Sentiment viz. Sentiment Viz is an online tool which can capture tweets of twitter by just entering the keywords of which we have to analyze the sentiments. It has different stylistic representation of the sentiments which include timeline, heatmap, tag cloud and many more. One more online tool that is used is “shared count”. Shared Count helps in analyzing the Facebook reactions, comments, shares done on a particular website.

## **Twitter API**

Application Programming Interface (API) allows two application to talk to each other. In the case of twitter API, this API lets an individual to read and write twitter data. This API is a licensed API means it can be used for the appropriate purpose only. This API allows the users to copy only reasonable and sensible amount of data. It has many rules, twitter policies, API policies which everyone has to follow in order to use this API. ([developer.twitter.com](http://developer.twitter.com), n.d.) In this paper, twitter API is used only for analyzing the sentiment of tweets made by public.

## Workflow of the above tools in this research



- Social media data like Twitter (tweets, comments and retweets), Facebook (Reactions, shares and comments), YouTube (number of videos posted, views on videos) of both Coca Cola and Pepsi are extracted using octoparse by copying and pasting the webpage link in Octoparse then making a workflow for extraction.
- Last year net income and revenue of both Coca cola Company and Pepsi Company are extracted using Google Chrome extension i.e., Data Miner and Scraper.
- Year wise net income and revenue of both the companies is extracted using Python with the help of Beautiful Soup API.
- All the extracted data then saved to excel file for cleaning (removing unnecessary data) and then combining meaningful sheets into one sheet.
- Then excel files used as a data source in Tableau for making specific graphs and dashboard and for differentiating between the two companies.
- Then Jupyter Notebook is used for investigate sentiment of twitter tweets of both the companies for a period of six days using Textblob and NLTK.

## Visualization and Results

The required data was collected manually from both the company respective social web pages. Let's have a look on Coca Cola and Pepsi parent company (i.e., Coca Cola Co and PepsiCo) Net income and Net revenue for the last 15 years. (See Appendix A8)

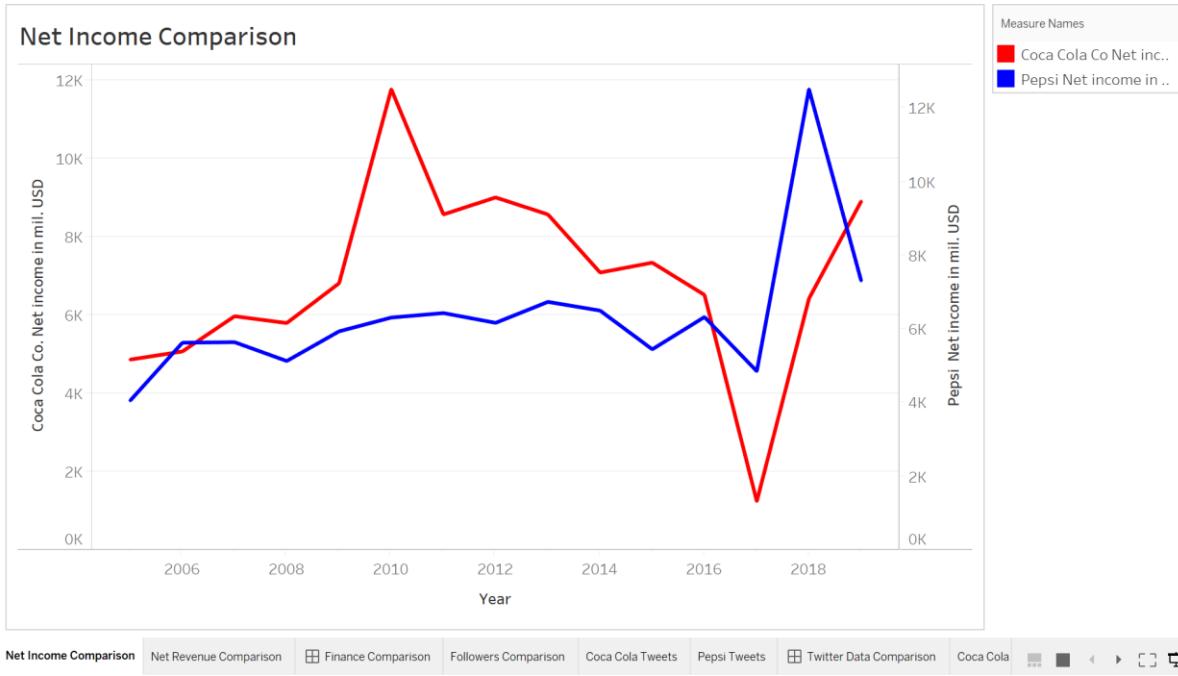


Fig 1



Fig 2

The following result from Fig 1 and Fig 2 shows that Net income of Coca Cola Co was higher than PepsiCo in many years but in 2017 the net income of Coca Cola Co goes much low as compared to previous years and in 2018 PepsiCo net income goes much high as compared to previous year. In the case of Net Revenue, PepsiCo is always ahead of Coca Cola Co.

In the case of PepsiCo, it owns many individual brands which includes food, snacks and drinks on the other hand Coca Cola Co only owns drinks. As in this paper we are only considering and comparing the beverages (i.e., Coca Cola and Pepsi) of both the companies. According to Investopedia, more than half of the revenue of PepsiCo comes from food and snacks only (Investopedia, 2019). That means PepsiCo approximate beverage revenue is half of the total revenue. See Table 1 for comparing both companies' revenue only for beverages.

PepsiCo Average Revenue for 15 years (Snacks + Food + Beverages) in mil USD	PepsiCo Average Revenue only for Beverage (total – 50%(approx.)) in Mil USD	Coca Cola Co Average Revenue for 15 years in Mil USD
55852	27926	36976

Table 1

The following table 1 shows the result that in terms of Beverages, Coca Cola has a high net revenue the Pepsi.

## Coca Cola vs Pepsi Followers

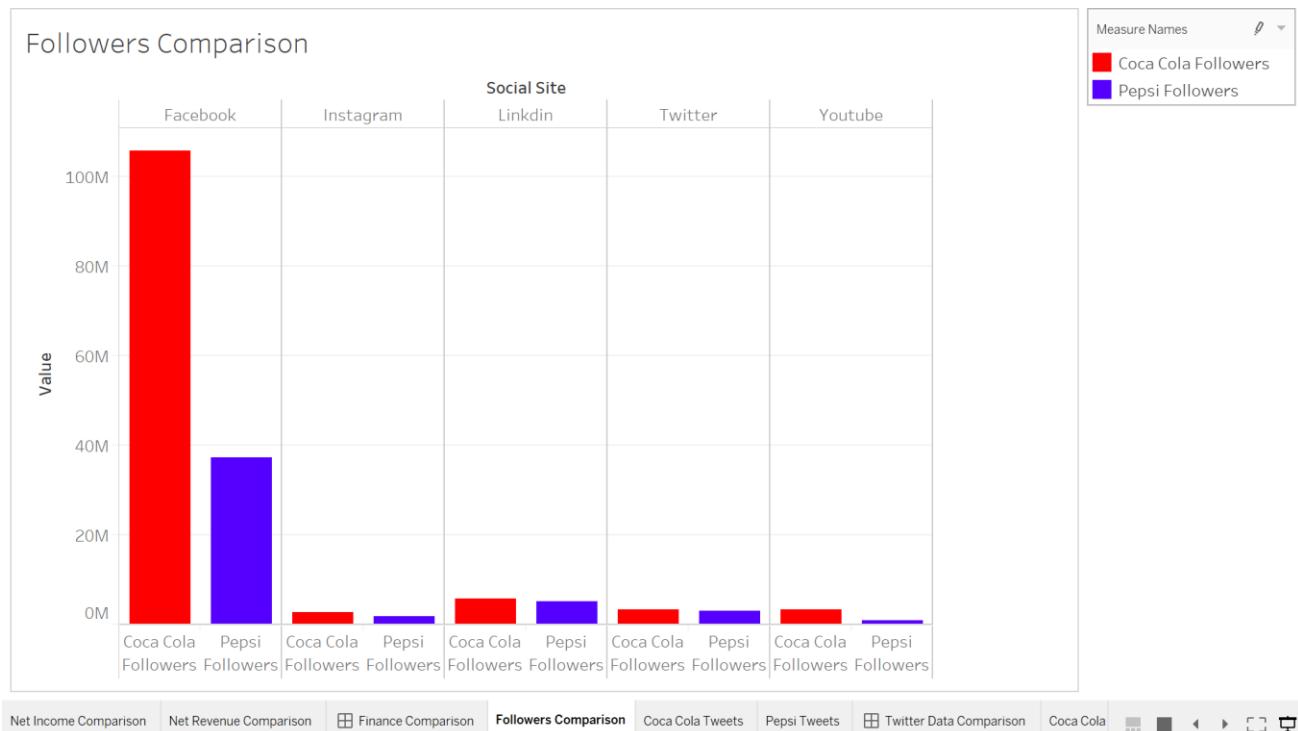


Fig 3

According to Fig 3, it is clearly showing that Coca Cola have more followers than Pepsi on the popular social media sites which includes Facebook, Instagram, LinkedIn, Twitter, and YouTube. This benefited Coca Cola in growing its brand value compared to Pepsi as Pepsi is not followed as much on social media. (See Appendix A8)

### Coca Cola Vs Pepsi on Twitter

Coca Cola is amazing force on twitter with 3.3 million followers on its main page which beats Pepsi main page with 3 million followers. Let's have a look on total number of tweets done by both the brands. See Fig 4.



Fig 4

It shows that Coca Cola tweets regularly its own content and also has a great strategy of replying to its followers' tweets on the other hand Pepsi has a very less engagement in terms of tweets as they tweet rarely than its competitor.

Now let's have a look on data extracted through twitter page. Fig 5 compares the extracted tweets from both Coca Cola and Pepsi twitter page.

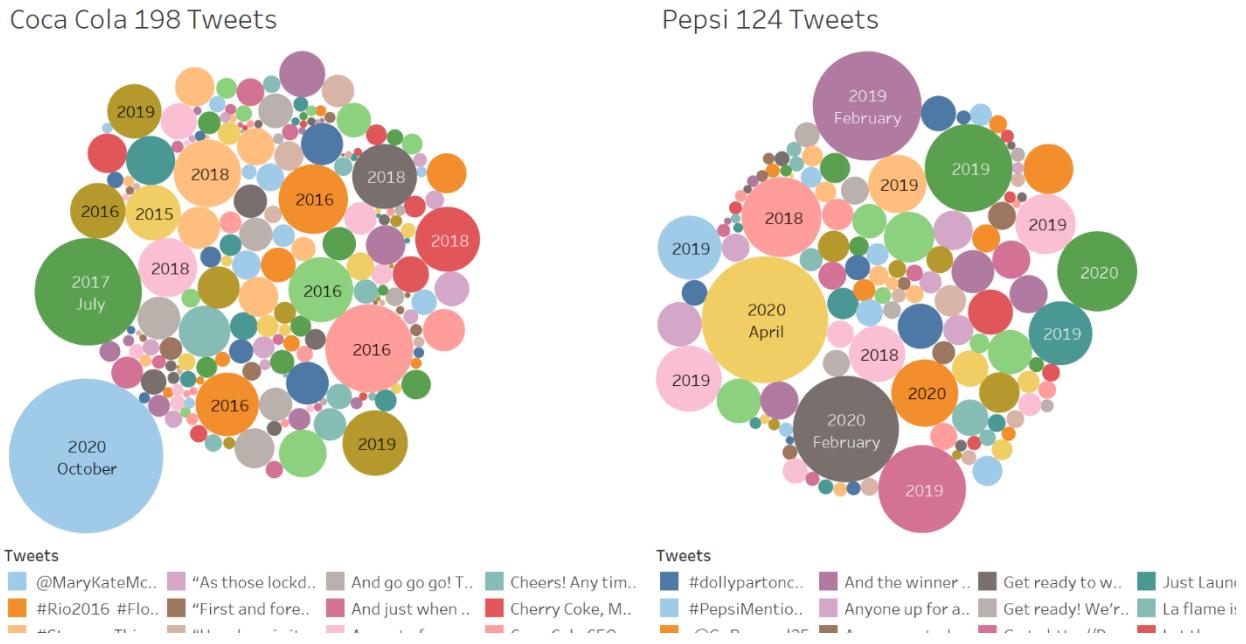


Fig 5

The comparison is between 198 coca cola tweets and 124 Pepsi tweets that was extracted in a same time of duration. I've used packed bubbles for differentiating both brands tweets. The bubbles contain the title, month and year of the tweet, the tweet itself and number of comments, likes and retweets on a particular tweet. The size of the bubbles indicates the number of likes in which Coca Cola and Pepsi got a maximum of 8900 and 2700 likes respectively on one of their tweets. When comparing this maximum like with the total followers of the page the Coca Cola tweet has a like to followers' ratio of 0.002 whereas Pepsi has the ratio of 0.0009. This means Coca Cola fans clicked the like button on average more times than Pepsi fans. For a clear view of this comparison refer Appendix A8.

Now that we know the most important data from twitter its time to focus on Facebook.

## Coca Cola vs Pepsi on Facebook

The biggest difference between the competitors we find is total number of followers. To be specific the Coca Cola Facebook page has 105 million followers and the Pepsi Facebook page has only 37 million followers, which means that the coca cola is 68 million ahead of Pepsi in terms of followers. (See Fig 3 and Appendix A8)

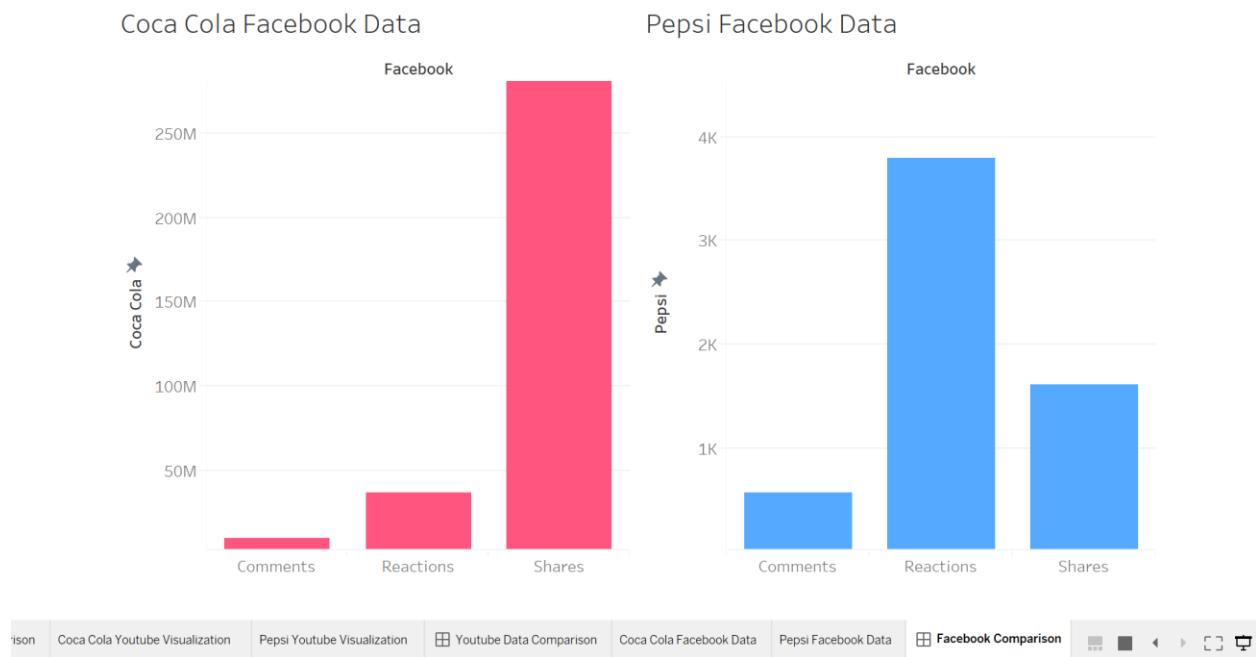


Fig 6

Fig 6 compares the interaction of peoples between all the Coca Cola and Pepsi social media sites with Facebook. It is clearly showing that Coca Cola rules in front of Pepsi with a huge difference. In all the three categories (i.e., Comments, Reactions and Shares) Coca Cola strikes against Pepsi. The biggest difference is between the Facebook shares. Pepsi has only 1617 shares whereas Coca Cola has around 286 million of shares. Although reactions and comments has also a very large difference.

## Coca Cola vs Pepsi on YouTube

Coca Cola has 34.7 lakh subscribers on its YouTube main page whereas Pepsi has only 8.53 lakh subscribers. Another huge difference between these two rivals. Again, Coca Cola strikes here with a massive lead. The total number of videos uploaded by both the Coca Cola and Pepsi are 3790 and 42 respectively. (See Fig 7)

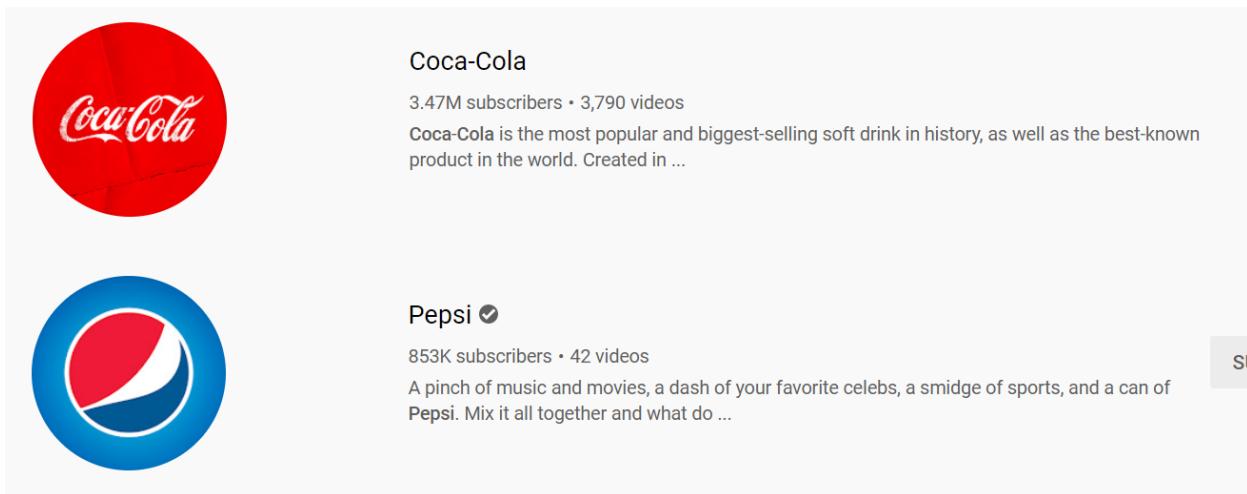


Fig 7

Now let's have a look on data extracted from YouTube main pages. Fig 8 compares the uploaded YouTube videos of both Pepsi and Coca Cola

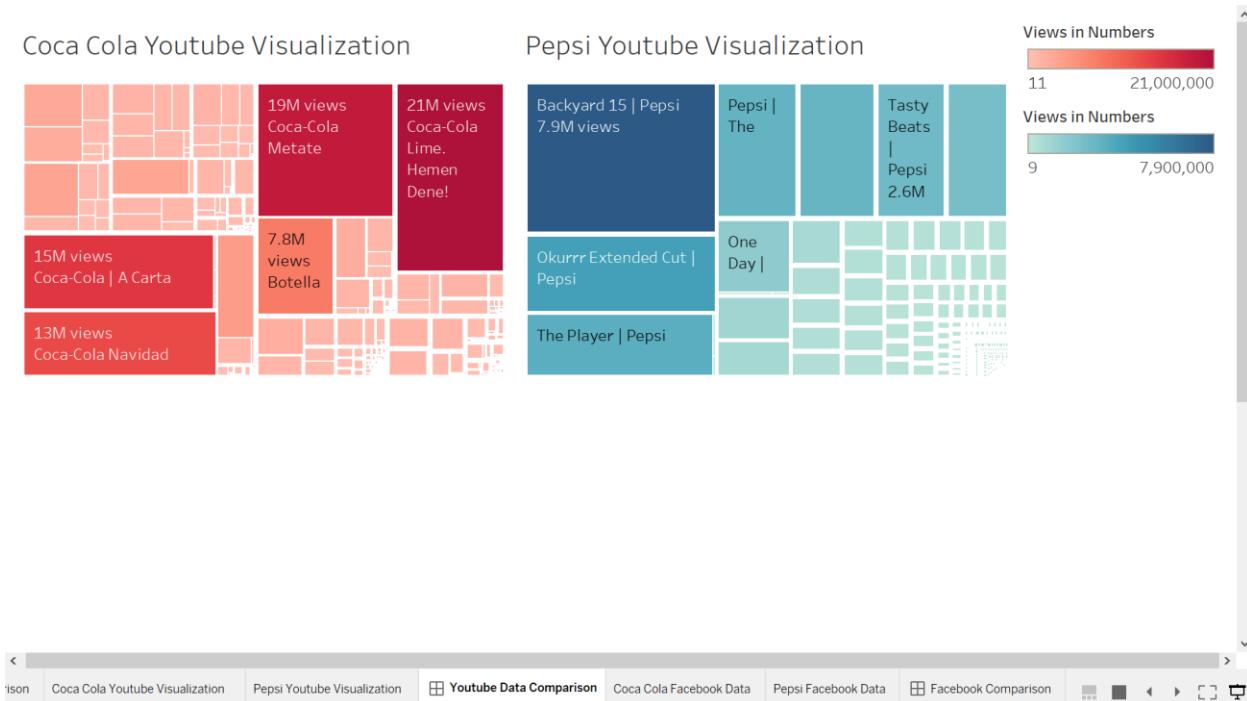


Fig 8

For comparing the YouTube videos I've used both parent company and the brand YouTube page (Pepsi, PepsiCo and Coca Cola, Coca Cola Co) by searching keyword (i.e., #pepsi and #cocacola) on their parent company YouTube page i.e., (PepsiCo and Coca Cola Co). This comparison is between 1049 Coca Cola YouTube videos and 305 Pepsi YouTube videos that was extracted in a same time interval. I've used Tree Maps for differentiating both brands YouTube data. The trees or we can simply say the rectangles contains the title, duration, uploaded time and the number of views of the particular videos. The size of the rectangles indicates the number of views. In this comparison Coca Cola got a maximum view of 21 million on one of their uploaded video where as Pepsi got 7.9 million of highest views on their video. For a clear view of this comparison refer Appendix A8.

### Sentiment Analysis

Sentiment Analysis is an interesting topic in todays generation in the fields of Machine Learning and Natural Language Processing. It is a process of examining a text, feeling or opinions.

In this research, I conduct a six days Sentiment Analysis on the tweets of both the brands to analyze sentiments on every particular day. I used two platforms for analyzing the sentiments.

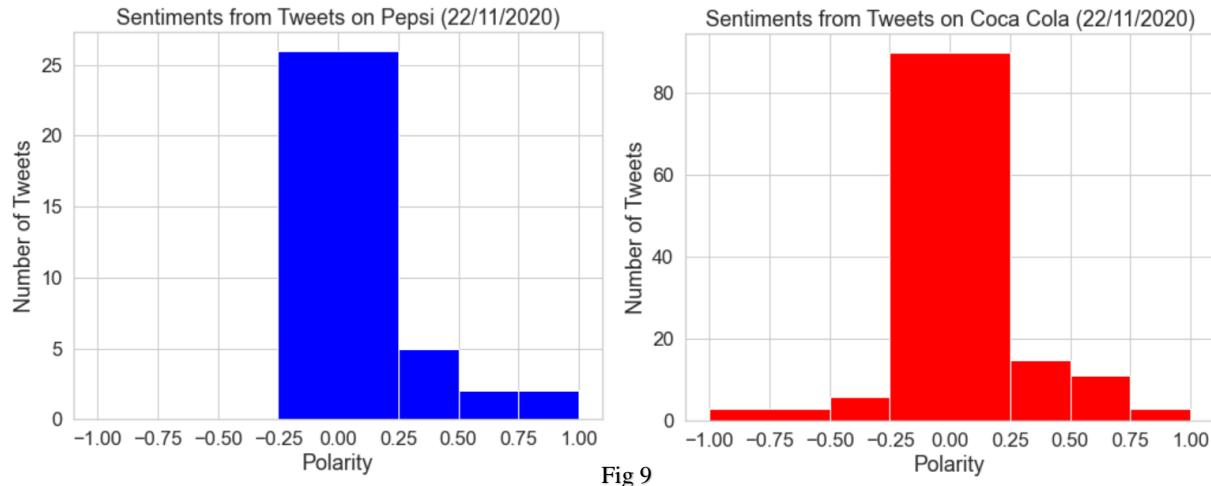


Fig 9

Fig 9 compares sentiment of both Coca Cola and Pepsi for one day using bar graph. The y-axis and x-axis of the graph relates with number of tweets and polarity respectively. I've categorized tweets with polarity greater than zero as "Positive", less than zero as "Negative" and zero as "Neutral". On this particular day Coca Cola have all "negative", "positive" and "neutral" tweets whereas Pepsi has only "positive" and "neutral" tweets. For all 6 days comparison See Appendix A9 and Appendix A10

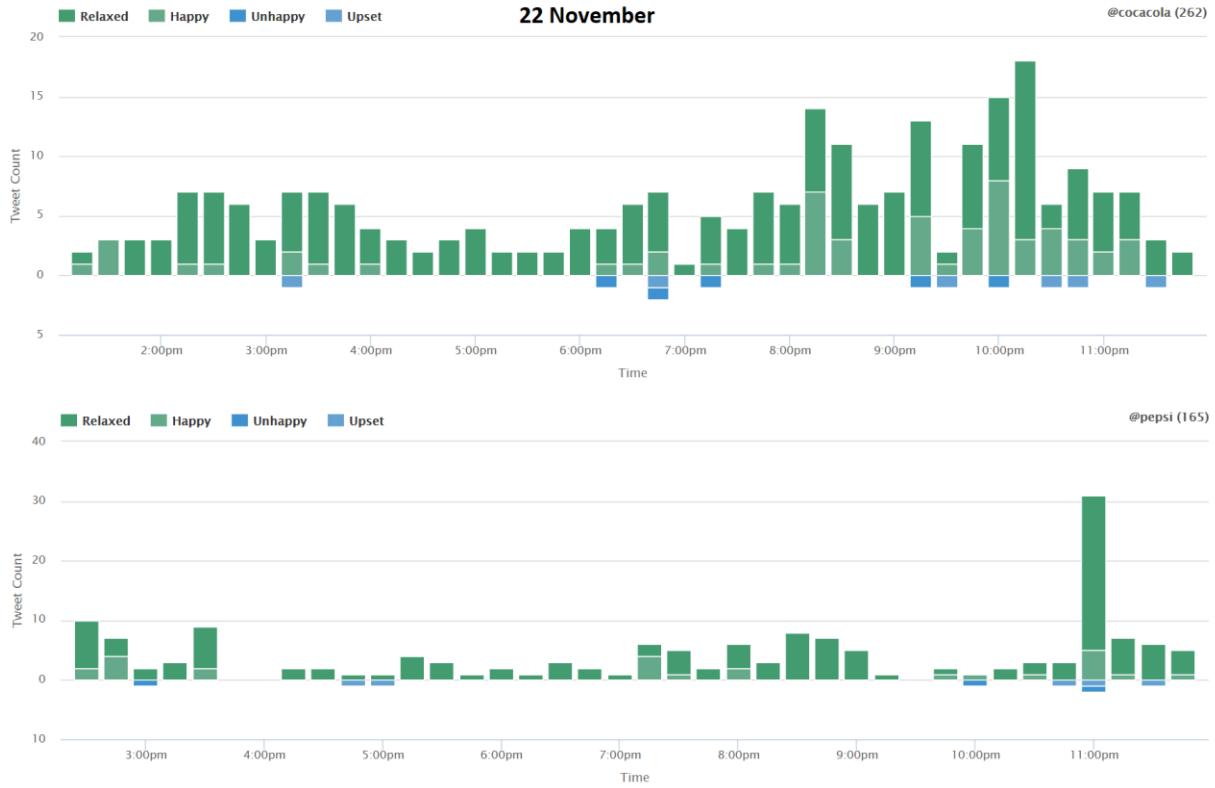


Fig 10

Fig 10 compares the one-day timeline of people emotions while tweeting for Pepsi and Coca Cola. 262 total tweets about Coca Cola on this particular day whereas 165 tweets for Pepsi. On comparing the emotions of the tweets less people are unhappy and upset for both Coca Cola and Pepsi but Coca Cola has higher happy and relaxed tweets as compared to Pepsi. From this Sentiment, we can conclude that on this particular day Coca Cola gets more and better review on twitter as compared to Pepsi. There is a possibility that Coca Cola is more active on twitter. For all six days comparison see Appendix A11, Appendix 12, Appendix 13, Appendix 14.

After comparing all the main social media for both Coca Cola and Pepsi it comes to an end that Coca Cola is the landslide winner with having a great strategy of focusing on all social media accounts and have more social media engagement as compared to Pepsi in almost all circumstances. Pepsi have to involve more and have to find some creative ideas to increase their engagement.

## **Limitation**

However, this research has several limitations –

1. The primary limitation of this research is that it analyses only for a specific period of time which means the results could change in the future.
2. Another limitation is that minimum amount of data is extracted through Octoparse due to which the data sets tends to smaller in size.
3. Sometimes duplicate data extracted through Octoparse which have to be cleaned further.
4. Computer algorithms have problems recognizing things like Jokes and Sarcasm in analyzing sentiments. Sometimes it excludes unrelated sentiment.
5. Sentiments algorithm cannot recognize hashtags.
6. Instagram data is hard to extract through Octoparse as the likes and comments of photos are on the photos itself (not below the photo) which the software is not able to recognize.
7. Facebook data is hard to extract.

## **Conclusion and Recommendation**

From the above results I have concluded that when it comes to Social Media, I'm getting nothing from Pepsi. I have no specific affinity for both the brands but at least Coca Cola has a consistency towards their brand as they are engaging their customers more on social media rather than in physical stores.

In today's time, people are more engaged on Twitter and Facebook as compared to other social media as they are trending more. So, it is recommended that the upcoming beverage brand should focus on the trending social media more. I would recommend the upcoming brand to take some ideas from Coca Cola social platforms to have a great success in the future.

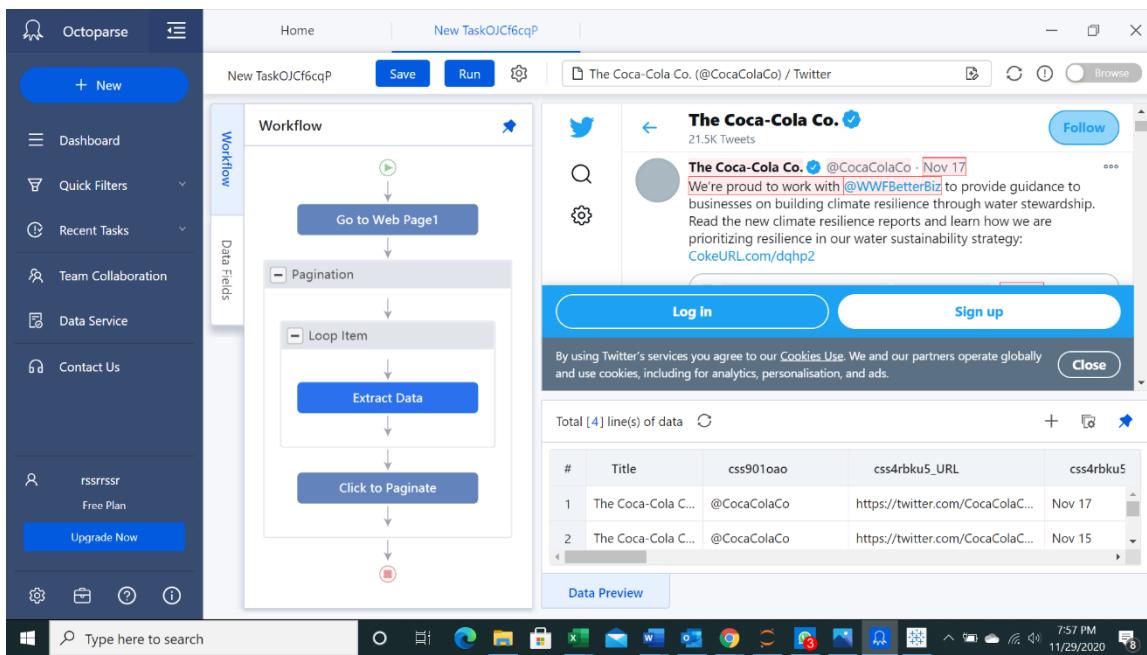
## References

1. Plume, C.J., Slade, E.L. and Dwivedi, Y.K. (2017). *Social media in the marketing context : a state of the art analysis and future directions*. Amsterdam Boston Heidelberg London New York Oxford Paris San Diego San Francisco Singapore Sydney Tokyo Elsevier Chandos Publishing.
2. Frankenfield, J. (2019). *What Is Social Data?* [online] Investopedia. Available at: <https://www.investopedia.com/terms/s/social-data.asp> [Accessed 25 Nov. 2020].
3. Wikipedia Contributors (2019a). *Coca-Cola*. [online] Wikipedia. Available at: <https://en.wikipedia.org/wiki/Coca-Cola> [Accessed 25 Nov. 2020].
4. Wikipedia Contributors (2019b). *Pepsi*. [online] Wikipedia. Available at: <https://en.wikipedia.org/wiki/Pepsi> [Accessed 25 Nov. 2020].
5. Batrinca, B. and Treleaven, P.C. (2014). Social media analytics: a survey of techniques, tools and platforms. *AI & SOCIETY*, 30(1), pp.89–116.
6. Team, O. (2020). *About Octoparse*. [online] www.octoparse.com. Available at: <https://www.octoparse.com/blog/what-is-octoparse> [Accessed 26 Nov. 2020].
7. Data Miner. (n.d.). *Extract data from any websites into a spreadsheet*. [online] Available at: <https://data-miner.io/how-it-works> [Accessed 27 Nov. 2020].
8. chrome.google.com. (n.d.). *Scraper*. [online] Available at: <https://chrome.google.com/webstore/detail/scraperv/mbigbapnjcgaffohmbkdlcacepngjd?hl=en> [Accessed 27 Nov. 2020].
9. How To Developers: Using Python for the Cloud and Big Data Analytics. (2015). [online] 1 Oct. Available at: <https://search-proquest-com.ezproxy.cardiffmet.ac.uk/docview/1719231441?accountid=15588&pq-orignsite=primo>.
10. Educative: Interactive Courses for Software Developers. (n.d.). *What is Spyder?* [online] Available at: <https://www.educative.io/edpresso/what-is-spyder>.
11. docs.python.org. (n.d.). *urllib.request — Extensible library for opening URLs — Python 3.8.6rc1 documentation*. [online] Available at: <https://docs.python.org/3/library/urllib.request.html> [Accessed 27 Nov. 2020].

12. Idris, I. (2016). *Python data analysis cookbook*. Packt Publishing Limited.
13. Richardson, L. (2019). *Beautiful Soup Documentation — Beautiful Soup 4.4.0 documentation*. [online] Crummy.com. Available at: <https://www.crummy.com/software/BeautifulSoup/bs4/doc/>.
14. Excel-easy.com. (2010). *Data Analysis in Excel*. [online] Available at: <https://www.excel-easy.com/data-analysis.html> [Accessed 28 Nov. 2020].
15. Tableau. (n.d.). *What is Tableau?* [online] Available at: <https://www.tableau.com/why-tableau/what-is-tableau> [Accessed 28 Nov. 2020].
16. By Jupyter It All Make Sense. (2018). [online] 1 Nov., pp.145–146. Available at: <https://media-nature-com.ezproxy.cardiffmet.ac.uk/original/magazine-assets/d41586-018-07196-1/d41586-018-07196-1.pdf>.
17. Mckinney, W. (2018). *Python for data analysis : data wrangling with pandas, NumPy, and IPython*. Sebastopol, Ca: O'reilly Media, Inc., October.
18. Garcia, M. (n.d.). *How to Make a Twitter Bot in Python With Tweepy – Real Python*. [online] realpython.com. Available at: <https://realpython.com/twitter-bot-python-tweepy/#:~:text=Tweepy%20is%20an%20open%20source> [Accessed 28 Nov. 2020].
19. Perkins, J. (2010). *Python text processing with NLTK 2.0 cookbook*. Birmingham ; Mumbai Packt Publishing.
20. Shubham Jain (2018). *Natural Language Processing for Beginners: Using TextBlob*. [online] Analytics Vidhya. Available at: <https://www.analyticsvidhya.com/blog/2018/02/natural-language-processing-for-beginners-using-textblob/> [Accessed 29 Nov. 2020].
21. developer.twitter.com. (n.d.). *Developer Agreement and Policy – Twitter Developers*. [online] Available at: <https://developer.twitter.com/en/developer-terms/agreement-and-policy> [Accessed 26 Nov. 2020].
22. Investopedia. (2019). *Comparing Coca-Cola and Pepsi's Business Models*. [online] Available at: <https://www.investopedia.com/articles/markets/081415/comparing-cocacola-and-pepsi-business-models.asp> [Accessed 28 Nov. 2020].

## Appendix

### Appendix A1. Octoparse running window



### Appendix A2. Scraper

The screenshot shows the Scraper tool interface for extracting data from The Coca-Cola Company's Wikipedia page. The configuration includes:

- Selector:** XPath //div[1]/table[1]/tbody/tr[td]
- Columns:** \*[1] Type and \*[2] Column 2
- Filters:** Exclude empty results

The data extracted is displayed in a table:

Type	Column 2
1 The Coca-Cola Company's headquarters in Atlanta	Public
2 Type	NYSE: KODIJA components&P 100 component&P 500 component
3 Traded as	US\$12.161007
4 ISIN	Beverage
5 Industry	January 29, 1892; 128 years ago (1892-01-29)Atlanta, Georgia, U.S.
6 Founded	John Stith Pemberton (as Coca-Cola) Asa Griggs Candler (as The Coca-Cola Company)
7 Founders	Atlanta, Georgia, United States
8 Headquarters	Worldwide
9 Area served	James Quincey (Chairman and CEO) Brian Smith (President and COO)
10 Key people	List of The Coca-Cola Company products
11 Products	US\$37.27 billion (2019)[1]
12 Revenue	US\$10.09 billion (2019)[1]
13 Operating income	US\$8.92 billion (2019)[1]
14 Net income	US\$96.38 billion (2019)[1]
15 Total assets	US\$18.98 billion (2019)[1]
16 Total equity	86,200 (2019)[1]
17 Number of employees	List of The Coca-Cola Company subsidiaries
18 Subsidiaries	www.coca-colacompany.com
19 Website	

## Appendix A3. Data Miner

The screenshot shows the Data Miner application interface. On the left, there's a sidebar with icons for Public, Bookmarks, and My recipes. The main area displays several recipe cards:

- Auto-generated Recipe (9cG39)**: Columns (2) : Type | Column 2 | Site: en.wikipedia.org
- Recipe 1**: Columns (1) : Column 1 |
- Recipe 1**: Columns (1) : Column 1 |
- Auto-generated Recipe (32Rn9)**: Columns (5) : Year | Revenue in mil. USD | Net income in mil. US |
- Recipe 1**: Columns (1) : Column 1 |
- Recipe 1**: Columns (1) : Column 1 |
- Auto-generated Recipe (x9R9R)**: Columns (1) : Text | Site: www.reddit.com

To the right, a large table titled "Auto-generated Recipe (9cG39)" shows extracted data for PepsiCo. The table has two columns: "Type" and "Column 2". The data includes:

Type	Column 2
PepsiCo's global headquarte...	
Type	Public
Traded as	NASDAQ: PEPNASDAQ-10...
Industry	BeveragesFood processing
Founded	August 28, 1920; 100 years ...
Founder	Caleb Bradham
Headquarters	Harrison, New York (in the h...
Area served	Worldwide
Key people	Ramon Laguarta(Chairman ...
Products	See list of PepsiCo products
Revenue	US\$67.16 billion (2019)[1]
Operating income	US\$10.29 billion (2019)[1]
Net income	US\$7.353 billion (2019)[1]
Total assets	US\$78.55 billion (2019)[1]
Total equity	US\$14.87 billion (2019)[1]
Number of employees	267,000 (2019)[1]
Subsidiaries	List of subsidiaries

At the bottom, it says "Current site: en.wikipedia.org". The taskbar at the bottom shows various open windows and the system clock.

## Appendix A4. Python – Spyder IDE

The screenshot shows the Spyder Python IDE interface. The top menu bar includes File, Edit, Search, Source, Run, Debug, Consoles, Projects, Tools, View, Help. The toolbar has various icons for file operations and navigation.

The code editor on the left shows a Python script named "coca-cola finance.py" with the following content:

```

1 # -*- coding: utf-8 -*-
2 """
3 Created on Sun Nov 22 03:55:03 2020
4
5 @author: ROHIT SAINI
6 """
7
8 import urllib.request
9 url = "https://en.wikipedia.org/wiki/The_Coca-Cola_Company"
10 page = urllib.request.urlopen(url)
11 from bs4 import BeautifulSoup
12 soup = BeautifulSoup(page, "Lxml")
13 print(soup.prettify())
14
15 all_tables=soup.find_all("table")
16 all_tables
17
18 right_table=soup.find('table', class_='wikitable float-left')
19 right_table
20
21 A=[]
22 B=[]
23 C=[]
24 D=[]
25 E=[]
26 for row in right_table.findAll('tr'):
27     cells=row.findAll('td')
28     if len(cells)==5:
29         A.append(cells[0].find(text=True))

```

The right side of the interface includes a "Source" tab, a "Console" tab with a help dialog, and a "Variable explorer" tab. The help dialog provides information on getting help for objects and activating automatic documentation. The console tab shows some initial output and configuration settings.

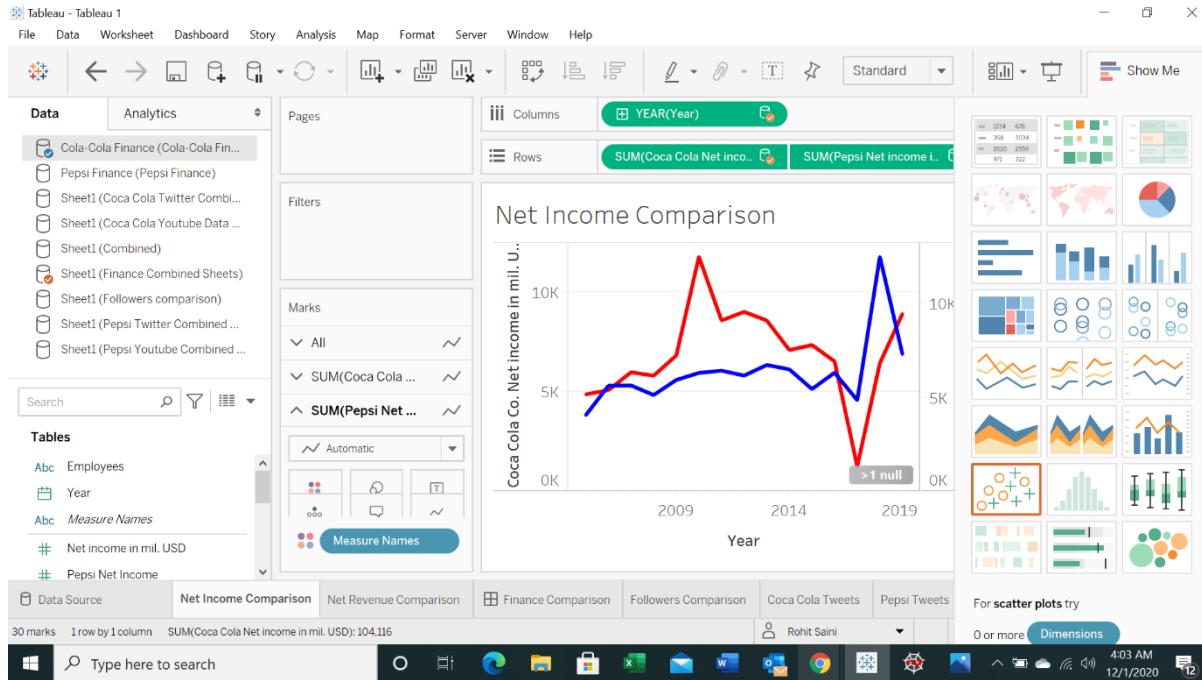
The taskbar at the bottom shows the operating system's taskbar with various icons and the system clock.

## Appendix A5. Excel

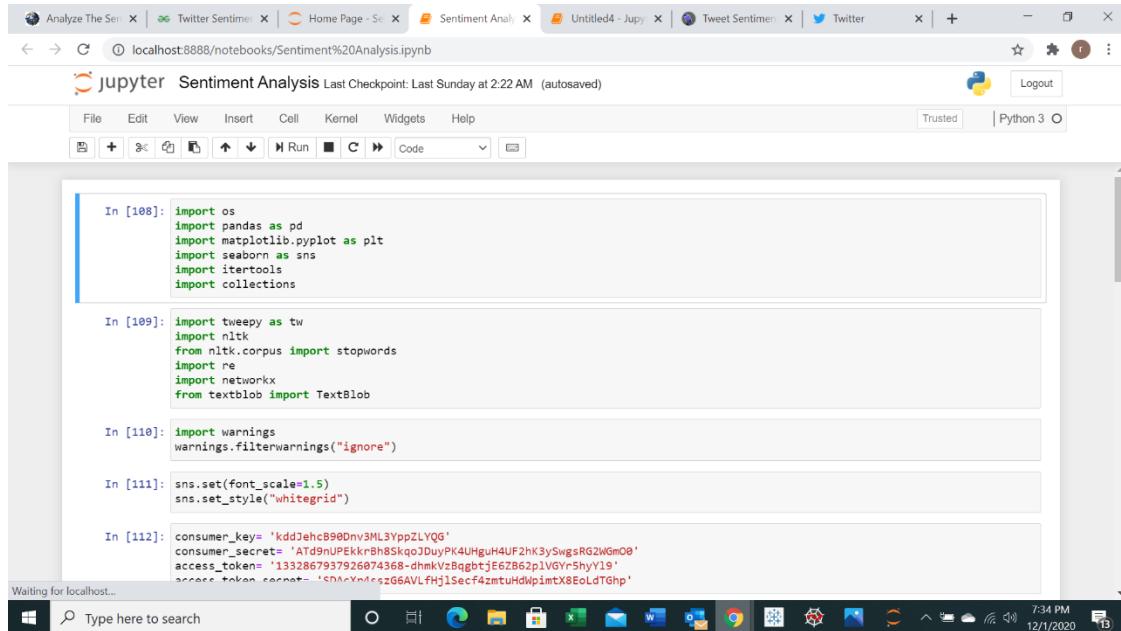
Coca Cola Youtube Data Combined - Excel

	A	B	C	D	E	F	G	H	I	J	K	L
1	Title	Score_URL	Image	Duration	Range	Views	Uploaded	Views in Numbers				
2	Julieta Martinez, líder de Tremen	https://www.youtube.com/watch?v=T...		1:00	NOW	PLA	11 views	44 minutes ago	11			
3	Eco-Lógica, gestora de residuos:	nhttps://www.youtube.com/watch?v=W...		1:00	NOW	PLA	14 views	1 hour ago	14			
4	Behavioural map - outside voice	rhttps://www.youtube.com/watch?v=ef...		9:00	NOW	PLA	51 views	23 hours ago	51			
5	Encouraging recycling behaviour	lhttps://www.youtube.com/watch?v=V...		18:13	NOW	PLA	48 views	1 day ago	48			
6	#UnMundoSinResiduos: sumamo	nhttps://www.youtube.com/watch?v=Rl...		1:49	NOW	PLA	272 views	4 days ago	272			
7	Fuerza, resistencia y solidaridad:	ihttps://www.youtube.com/watch?v=7e...		1:00	NOW	PLA	160 views	5 days ago	160			
8	Optimizar recursos	https://www.youtube.com/watch?v=9...		1:25	NOW	PLA	404 views	5 days ago	404			
9	Delivery	https://www.youtube.com/watch?v=oi...		1:42	NOW	PLA	354 views	5 days ago	354			
10	Mi salud y mi negocio	https://www.youtube.com/watch?v=b...		1:52	NOW	PLA	531 views	5 days ago	531			
11	María Ángelica, la almacenera que	https://www.youtube.com/watch?v=R...		1:00	NOW	PLA	521 views	1 week ago	521			
12	"Jardín secreto", una familia unida	https://www.youtube.com/watch?v=1j...		0:54	NOW	PLA	343 views	1 week ago	343			
13	Celebramos los 10 años de Sby20	https://www.youtube.com/watch?v=C...		1:01	NOW	PLA	739 views	1 week ago	739			
14	Celebramos los 10 años de Sby20	https://www.youtube.com/watch?v=V...		1:01	NOW	PLA	674 views	1 week ago	674			
15	Sby20 Compañía Coca-Cola- Mujer	https://www.youtube.com/watch?v=p...		2:44	NOW	PLA	2.5K views	1 week ago	2500			
16	Sby20 Compañía Coca-Cola- 5 mil	https://www.youtube.com/watch?v=O...		0:47	NOW	PLA	2.3K views	1 week ago	2300			
17	Webinar Las Mujeres Transforma	https://www.youtube.com/watch?v=2...		1:13:38	NOW	PLA	1.3K views	1 week ago	1300			
18	Estudiantes de ProgramON comp	https://www.youtube.com/watch?v=X...		0:57	NOW	PLA	421 views	1 week ago	421			

## Appendix A6. Tableau



## Appendix A7. Jupyter Notebook



The screenshot shows a Jupyter Notebook interface running in a browser window. The title bar indicates the URL is `localhost:8888/notebooks/Sentiment%20Analysis.ipynb`. The notebook contains several code cells:

- In [108]:

```
import os
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import itertools
import collections
```
- In [109]:

```
import tweepy as tw
import nltk
from nltk.corpus import stopwords
import re
import networkx
from textblob import TextBlob
```
- In [110]:

```
import warnings
warnings.filterwarnings("ignore")
```
- In [111]:

```
sns.set(font_scale=1.5)
sns.set_style("whitegrid")
```
- In [112]:

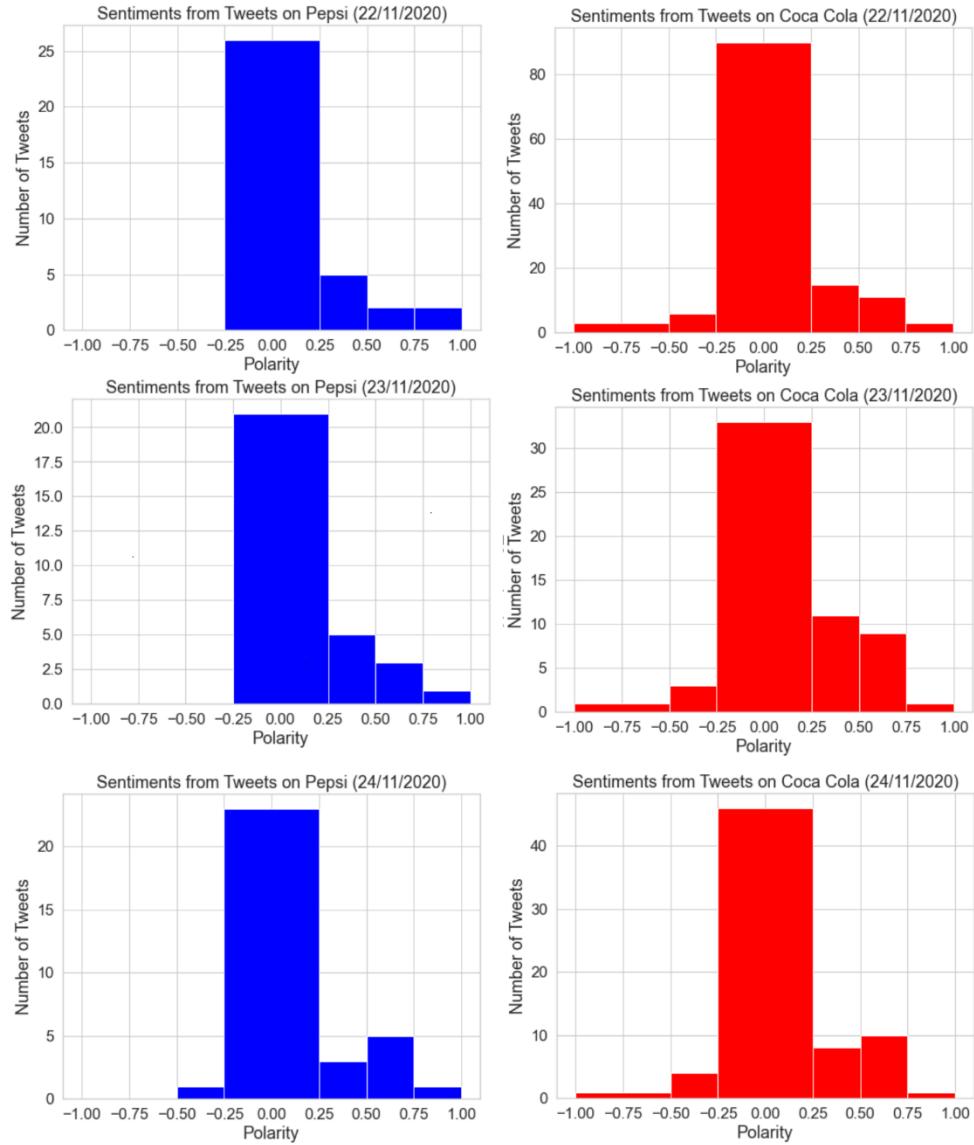
```
consumer_key='kddJehcB90Dnv3ML3YppZLYQ8'
consumer_secret='ATd9nUPekkrBh85kxqoJDuyPK4UHguH4UF2hK3ySwgsRG2WGm08'
access_token='1332867937926074368-dhnkvVzbqgbtjE62z862p1VGy-Shy19'
access_token_secret='eDAXXAmesG6AVLfhj1Secf4zmtuHdp1mtX8EoLdTchp'
```

The bottom of the screen shows a Windows taskbar with various icons and a search bar.

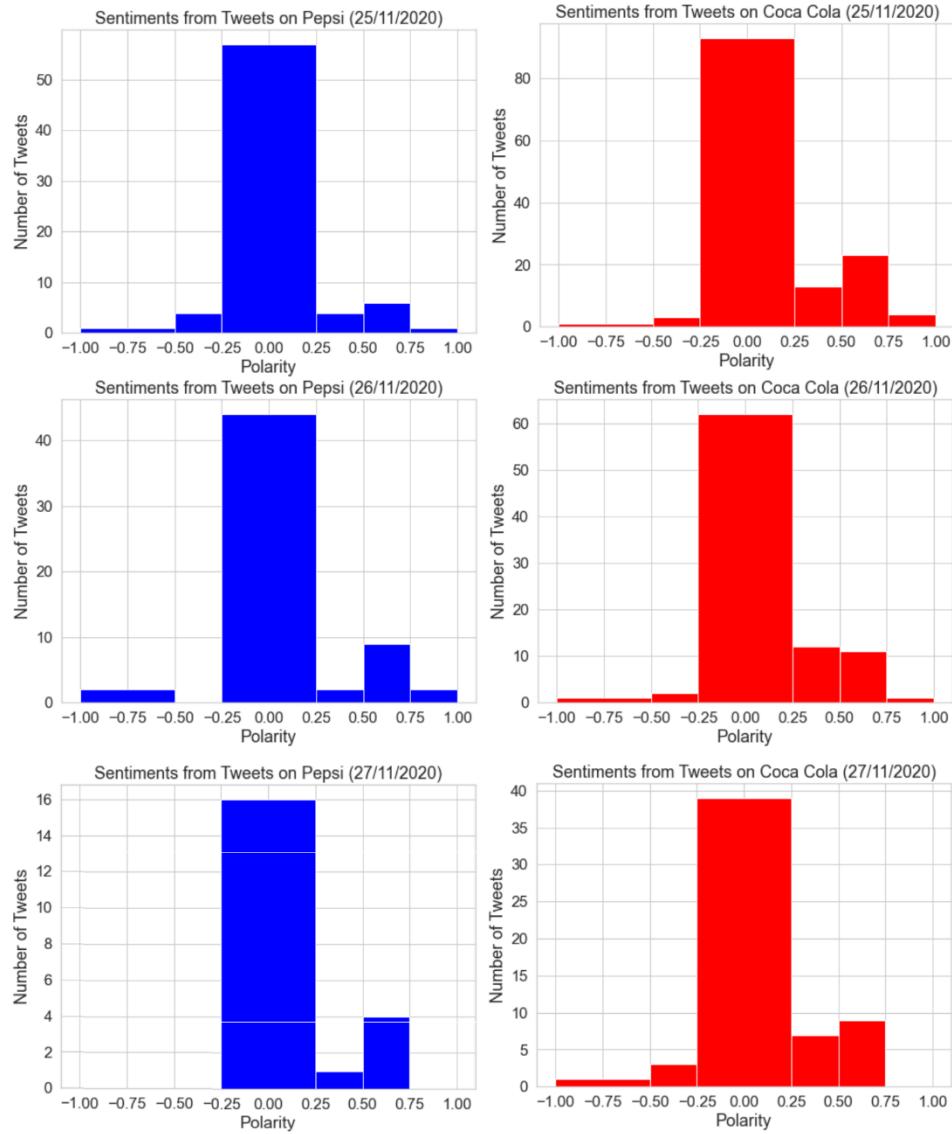
## Appendix A8. Tableau Public Link

<https://public.tableau.com/profile/rohit.saini3138#/vizhome/CocaColaVsPepsiSocialMediaAnalytics/NetIncomeComparison>

## Appendix A9. First three days Sentiment Comparison of Coca Cola and Pepsi Tweets



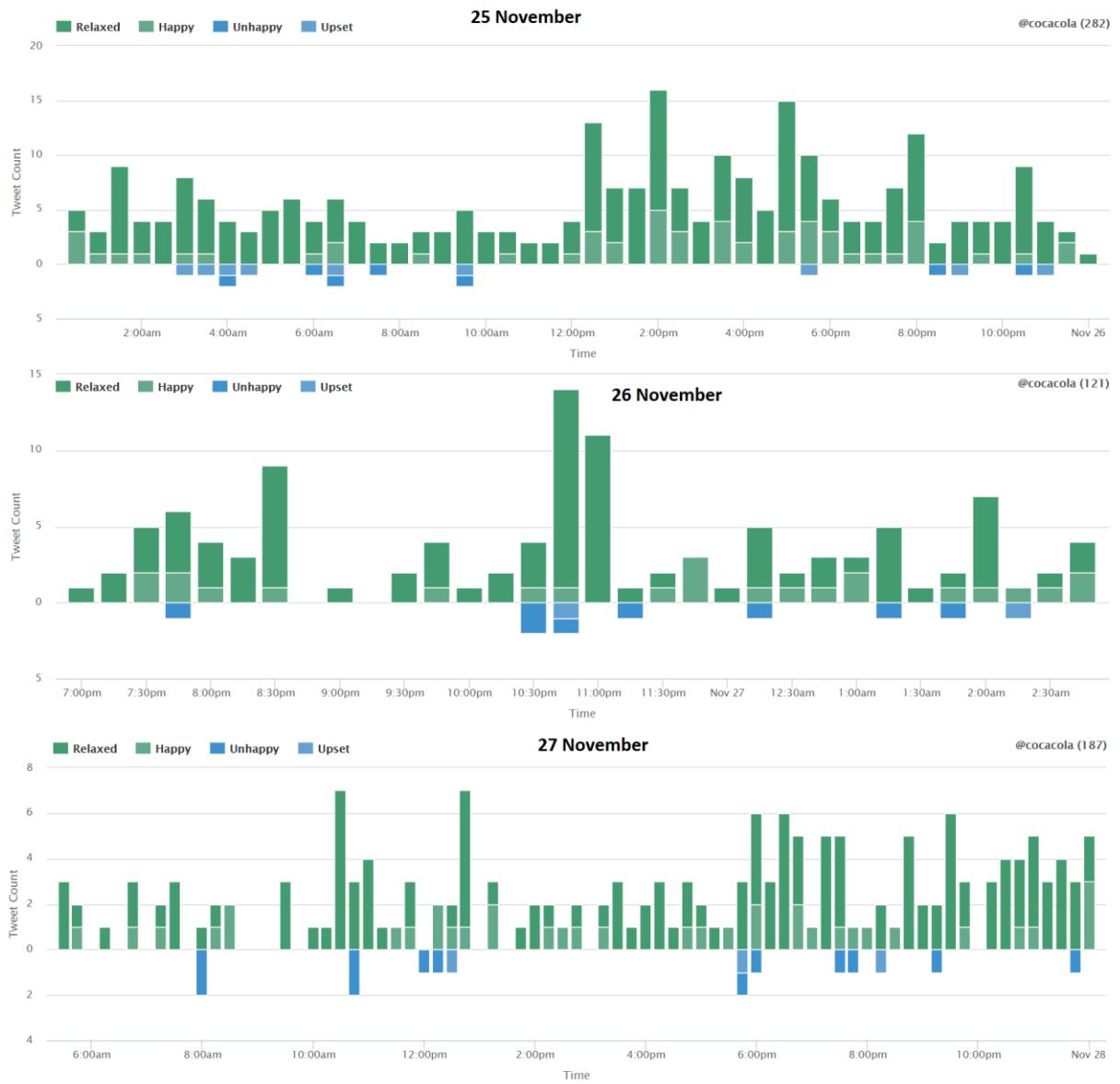
## Appendix A10. Last three days Sentiment Comparison of Coca Cola and Pepsi Tweets



## Appendix A11. Coca Cola First three days sentiment using Sentiment Viz



## Appendix A12. Coca Cola last three days sentiment using Sentiment Viz



## Appendix A13. Pepsi first three days sentiment using Sentiment Viz



## Appendix A14. Pepsi last three days sentiment using Sentiment Viz

