**1.** **Biography**

I am Md Emdad Hussain Imu. I have successfully completed my B.Sc in Computer science and engineering from Leading university. I have skill in C/C++, Java, Python, android app development, HTML and css. I also have some solid experience with data analytics, which would be a great addition to this course. I believe I am devoted to and passionate about finishing my master's in Data Science. I can be a valuable contributor to any organisation with my strong technological abilities and understanding.

**2. Abstract**

**2.1 Purpose**: From the start of the pandemic the whole world brought an end to the SARS-CoV-2. Pfizer also developed a vaccine for SARS-CoV-2. However, they become very popular in social media like twitter and gain a huge popularity among the people and it helps their business as well.

**2.2 Materials and methods:** For collecting data from twitter Octoperse application has been used. Sentiment analysis was performed using VADER (Valence Aware Dictionary for Sentiment Reasoning) to calculate the five levels of sentiments - overly positive, positive, neutral, negative, and overly negative.

**2.3 Finding:** Total 2020 tween has been used in this analysis. The sentiment regarding Pfizer remains positive and stable throughout the year.

**2.4 Business implications:** As the sentiment regarding about Pfizer remain positive throughout the time, Pfizer use this to capture huge market earn more revenue

**3. Introduction**

SARS-CoV-2 was formally classified as a public health emergency of global significance by the WHO on January 30, 2020. More than 140 million cases and 3 million fatalities had been reported globally as of 18 April 2021, (www.who.int, n.d.). The Pfizer/BioNTech vaccine was the first to receive approval for widespread use and was approved for use on December 2, 2020, just over a year after the pandemic was announced in the UK. In a remarkably short amount of time, a number of vaccinations have been created and licensed to combat the pandemic. The ChAdOx1 vaccine from AstraZeneca/Oxford, as well as messenger RNA (mRNA) vaccines developed by Pfizer/BioNTech and Moderna, are the primary vaccinations used in the Western world, despite the fact that a number of vaccines have currently been licensed globally.

In comparison to the other Covid-19 pharmaceutical companies over the previous two years, Pfizer received the highest mentions on Twitter and Reddit, according to Global Data’s Social Media Analytics database. In Q2 2021, the business posted a total of 9123 times on both social media platforms, which was a record high. Even though Moderna attracted the most attention on Twitter in the first half of 2020, Pfizer dominated the conversation by the end of the year due to the flurry of approvals for its mRNA Covid-19 vaccine. In the first half of 2021, Reddit saw an even greater surge in this chatter. Although its vaccine partner BioNTech experienced a similar trendline of peaked interest, the business did not achieve the same levels of notoriety.

**4. Design and discussion of data sources**

**4.1 Tweet retrieval**

For any data scientist, whether they are professionals, students, or amateurs, Twitter is a veritable gold mine. It links people from many walks of life, including local officials and previous and present presidents of the United States, as well as aspiring musicians like Taylor Swift and basketball players like LeBron James and rec league basketball players like LeBron.

How many people use it, and how much data does it produce? About 145 million individuals were daily active users of Twitter as of Q3 2019, and in the year of 2018, there were 500 million tweets sent out each day. It is not surprising that Twitter data offers a rich supply from which data insights can be gleaned, even with its 280-character constraint. (Hwang, 2020).

**4.1.1 Tweepy**

The open-source Python library Tweepy provides a very practical way to use Python to access the Twitter API. In addition to transparently handling different implementation details like data encoding and decoding, HTTP queries, results pagination, OAuth authentication, Rate restrictions, and Streams, Tweepy contains a set of classes and methods that represent Twitter's models and API endpoints.

If Tweepy was not used than the user would have to deal with low-level aspects relating to HTTP requests, data serialization, authentication, and rate constraints. This could be time-consuming and error-prone. Thanks to Tweepy, one may instead focus on the functionality they want to build.

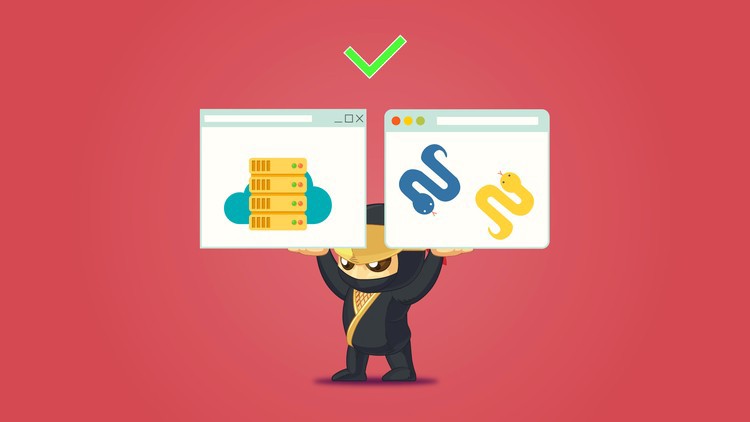
Access to almost all of the Twitter API's functionalities is made possible through Tweepy. The sole limitation as of version 3.7.0 is that due to some recent updates to the Twitter API, Direct Messages don't work correctly.

Tweepy provides access to the Twitter API for Python users. It accomplishes this by encapsulating a significant chunk of the Twitter API's complexity and adding a model layer and other useful functionalities on top of it.

Tweepy still uses some of the older phrases since the names of important Twitter ideas have evolved over time. as an illustration, a status update is a tweet, a friendship is a relationship between followers, and a favourite is a like. The following categories can be used to categorise Tweepy's functionality: OAuth, The API Class, Models, Cursors, and Streams.

Even though the Tweepy procedure is simple, a free Twitter developer account only allows us to go so far. We only have 7-day search history and receive 50 queries every month. (Python, n.d.).

**4.1.2 Selenium**

An open-source testing framework called Selenium Online Driver automates browsers to test web applications. It has the ability to test numerous browsers simultaneously. It facilitates testing across several browsers. It functions on a variety of operating systems. There is no specific server required to conduct tests using the Selenium Web driver. Additionally, the Web driver launches and manages a browser by itself.

The Selenium Webdriver interface offers a number of techniques, including; getTitle(), close(), quit(), etc. [Selenium online training](https://onlineitguru.com/selenium-training.html) helps to enter content in the way of structures.

Any web browser is interacted with by the Selenium web driver in a human-like manner. Additionally, the extremely quick HTML Unit browser is supported.

A high-level object-oriented programming language used for many applications is called Python. It includes a multipurpose coding language. Compared to other programming languages, it makes it easier for developers to read and translate Python code. It is crucial to the automation process. It makes the functions legible by using a straightforward syntax. Its stateless operations also prevent interruptions. The ability to reuse the same code across other projects is also helpful to developers.

**4.1.2.1 Selenium Web driver architecture**

This architecture consists of a few significant parts. These elements support web driver in a variety of ways. Additionally, this architecture demonstrates how a Selenium web driver operates. These include the Selenium Client Library, the JSON wire protocol, browsers, and browser drivers.

More quickly than any other Selenium suite, Selenium web driver aids in the execution of test scripts. Without a server needed, it connects directly with the browser. Additionally, it handles dynamic web components like notifications, checkboxes, etc. better. It has a collection of locators that make it easier to find web components on any web page. (Prasanna, 2020).

**Advantage**

As a free automated testing tool, it works with multiple operating systems and browsers. Since it is an upgraded version of RC, its performance must be quicker. It combines with various testing frameworks quickly. Additionally, it coordinates with every item.

**Cons**

It lacks an IDE for script creation. It makes use of other IDEs for scripting. It only supports web apps.

**4.1.3 Octoparse**

Octoparse is a strong and free website crawler that can extract nearly any type of data from a website. With its numerous features and operations, Octoparse can be used to rip websites. You can copy all of the website's text using the point-and-click interface. The term "Octoparse Cloud Extraction" describes a distributed computing 7/24 procedure used to get data on a huge scale from numerous cloud servers. After installing the app, you may create a new job, set up a workflow or rule for it, and then run it using cloud extraction. After that, you can shut down your computer and give Octoparse the job. With Scheduled Cloud Extraction, which is more complex, users can refresh the website and retrieve the most recent information from it. Using cloud services, users can accelerate their processes and set a certain time for them to run. Users may require a more organised and condensed data result format. Users could make use of Octoparse's RegEx Tool to standardise data fields and eliminate redundant regular expressions. By creating Regular Expression automatically based on users' formatting needs, this built-in utility will assist users in reformatting the retrieved data. Since users could accurately pinpoint web parts using the XPath configuration tool offered by Octoparse, customers were able to extract numerous challenging websites with complex data block layout utilising its built-in Regex and XPath tool. Since Octoparse offers IP Proxy Servers that automate IP's departure without being noticed by aggressive websites, users won't have to worry about IP banning any longer. After data extraction is complete, users can download almost the entire website and save it in an organised format like EXCEL, TXT, HTML, or any databases. Retrieved data can be used when the data scrape is finished for a variety of analyses, such as calculating the percentage of various sorts of art works.

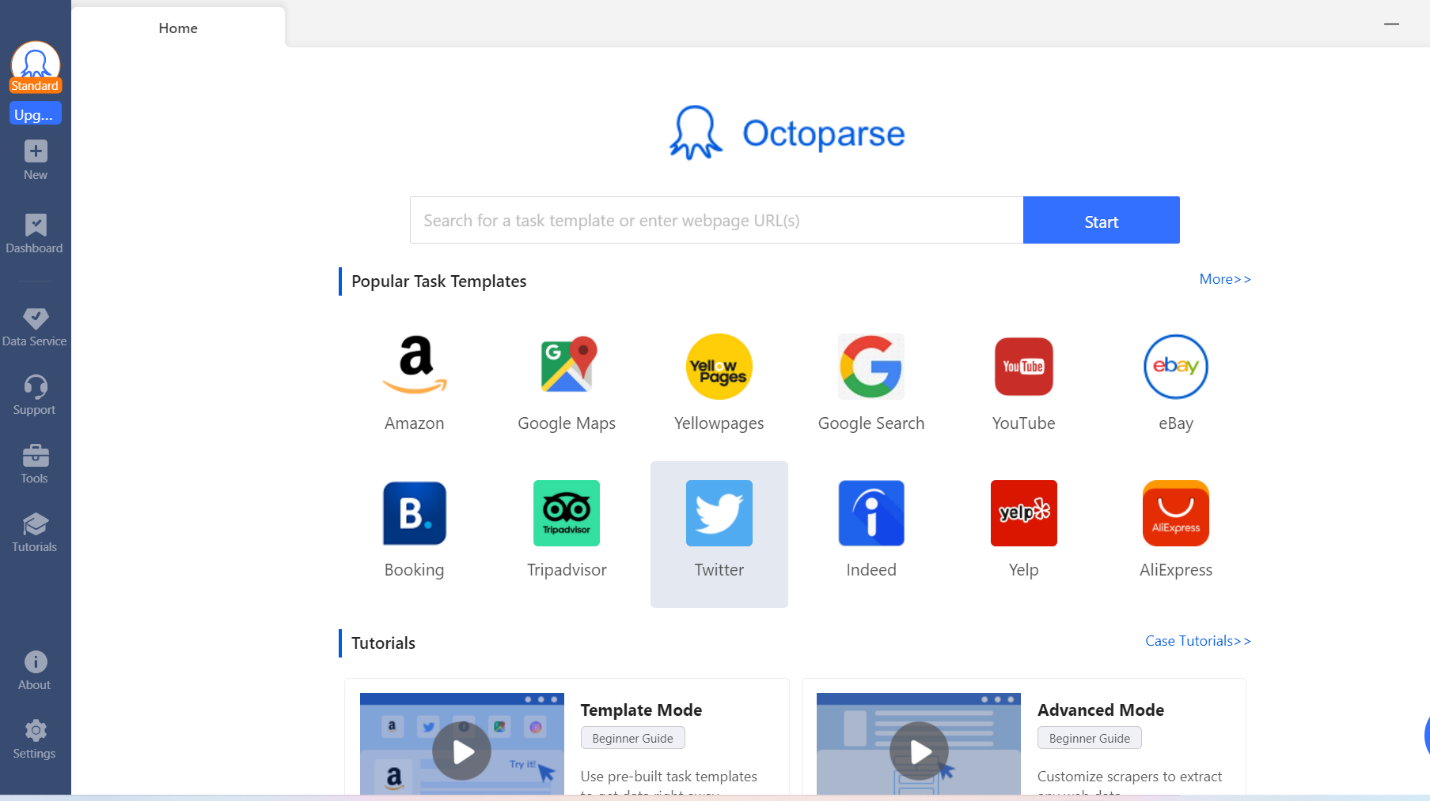


Figure 1: Interface of Octoparse

Whether using local or cloud extraction, Octoparse does not place any restrictions on the data that can be extracted. As much data as the website or task enables for extraction is available for extraction.

There is a restriction on free accounts for data export. A free account is limited to exporting 10,000 data rows at a time.

If one wants to export more data than 10,000 rows, they can upgrade their account to any paid plan or pay $5.9 for a one-time export credit. (Octoparse.com, 2019), (www.octoparse.com, n.d.).

**4.2 Sentiment analysis**

There are various types of analysis for sentiment analysis, which is a supervised machine learning task. Five levels of sentiments were taken into consideration: overly positive, positive, neutral, negative, and overly negative. A rule-based approach called VADER (Valence Aware Dictionary for Sentiment Reasoning) has been used to investigate how Pfizer's actions affected Twitter users' attitudes throughout the epidemic. The opinions expressed in a text can be categorized using sentiment analysis; the most frequent categories are "Positive," "Negative," and less commonly "Neutral." Through the use of machine learning algorithms and a classifier-based methodology, sentiment analysis can be completed. (Conner, et al., 2020)

**4.3 Justification of Choice**

A popular method for obtaining and extracting data from webpages is called web scraping. To accomplish their purpose, people start creating or utilising a number of different applications. They are typically split into the code and tools groups. In contrast to scraping with Python, Octoparse only requires entering the URL; applying for API is not necessary.

Compared to Python, Octoparse is easier to learn. It needs to grasp the web structure well and be aware of anti-scraping strategies in order to construct a Python crawler. This goes beyond simply being familiar with various libraries and coding techniques. But with Octoparse, all such scenarios have already been thought of for you, so all you need to do is click a few times to get all the information you need.

In this paper Octoparse is used because the data generated using python libraries (Selenium and Tweepy) that are not enough. It is inappropriate to use a free Twitter developer account for this paper because searches can only go back seven days and there are a maximum of 50 queries per month. Data scraping is simple with Octoperse.

**4.4 Visualization**

**4.4.1 Tablaue**

Tablaue has been used for visualization purposes. The business intelligence sector uses Tableau, a highly effective and quickly growing data visualisation application. It helps to simplify raw data into a very comprehensible format. Professionals at all levels of a business can understand the data thanks to Tableau's assistance in its creation. Users without technical expertise can also design unique dashboards. The Tableau programme quickly analyses data and produces dashboards and worksheets for visualisations. The wonderful thing about Tableau software is that it can be used by anyone without the need for technical or programming expertise. People from all spheres, including business, researchers, and different industries, have expressed interest in the instrument. Tableau requires much less time to create graphs than programmes like Microsoft Excel. Additionally, unlike Excel, Tableau employs algorithms to intelligently format graphs, utilizing most of the labour-intensive aspects of data processing. Excel files, text files, and even tables from PDF documents from which Tableau can extract the data are just a few of the formats that Tableau makes available for working with your raw data. Tableau may also connect to a database for an organization. The Data Interpreter is a feature in Tableau. Use it to tidy up imported Tables and get rid of any extraneous data that isn't relevant to the dataset. Tableau is reliable and doesn't crash while importing vast amounts of data, but Excel can't manage a lot of data. Tableau functions as an all-in-one solution, so no additional software is required to produce data presentations. It's entirely integrated in Tableau. Tableau advances by enabling the creation of dashboards. The fact that the data is interactive, the dashboard is controllable, and the display refreshes automatically to demonstrate that a query has been made for the data is arguably one of the most crucial aspects of how Tableau interacts with the dashboard. Despite the fact that Excel and other programs have begun to include more visualization features in their programs, Tableau continues to be the industry leader in data visualization tools, and users should switch to Tableau in order to increase productivity and gain deeper insights into the data. (www.octoparse.com, n.d.).

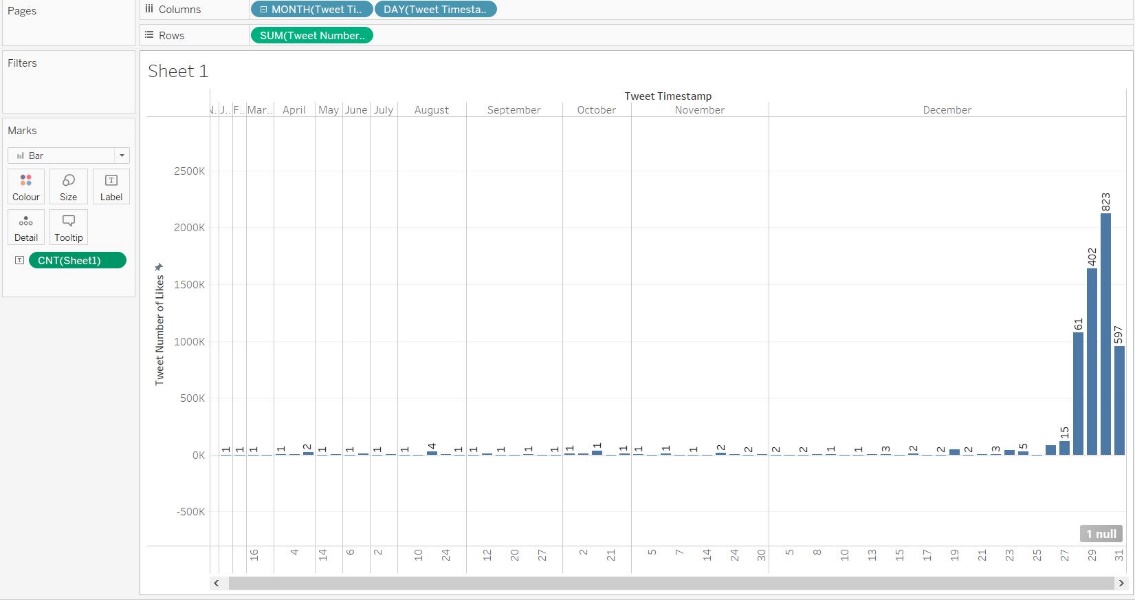
**4.5 Dataset description**

Dataset has been generated by Octoparse by scrapping twitter. The dataset consists of 2022 tweets that have been publish in year of 2021. After that, using python pre-processing step is done to make text lowercase, avoid punctuation emoji’s and stop-words. Then, to annotate the tweet as positive, negative or neutral the labelling step is applied. (Mee et al., 2021).

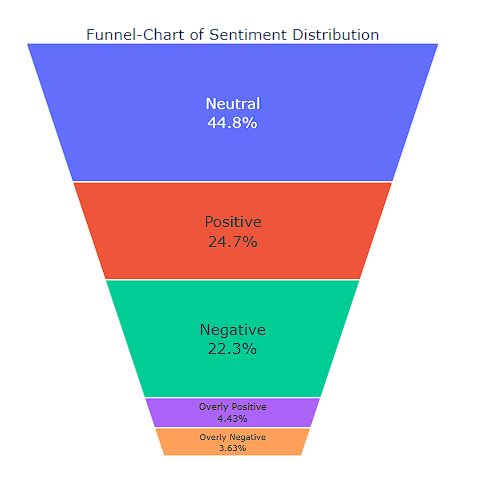
**5. Result**

**5.1 Visualizations**

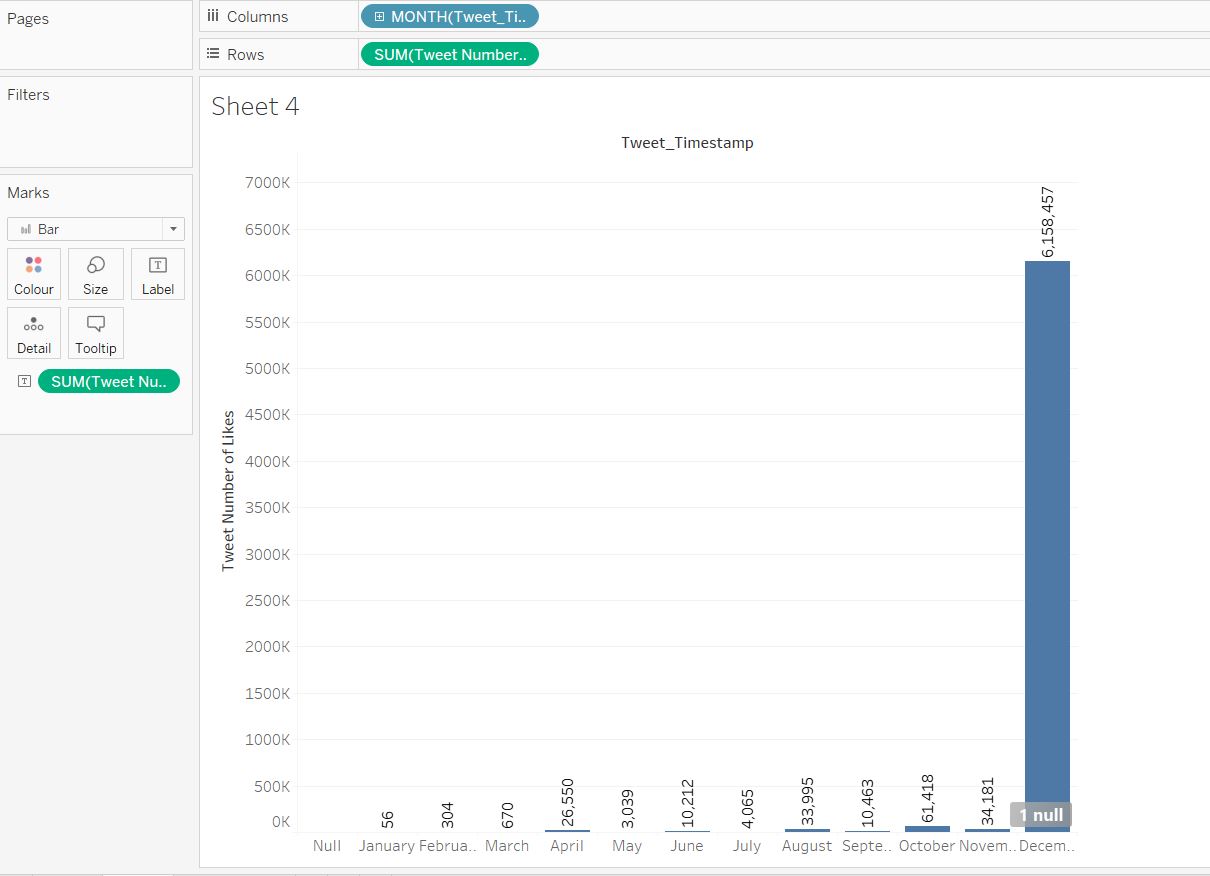
In the figure 2 it visualized that most people start to talk about Pfizer, this is the month when they declare that Children under the age of 5 are tested with a third dose of 3 µg, given at least two months after the second dose, and children aged 5 to under 12 are tested with a third dose of the 10-µg formulation.  (www.pfizer.com, n.d.). VADER divided the tweet into five category these are positive, overly positive, negative, overly negative and neutral. The majority of tweets (44.8%) fell into the neutral group, which was followed by the positive (24%), and the negative (25%) categories. In figure 6 we see that most of the people are more neutral or positive about the Pfizer vaccine. Positive publicity always helps a company to gain popularity.

****

**Figure 2:** Number of people talk about Pfizer in each month



**Figure 3**: Sentiment from the data



**Figure 4:** Number of like in post every month

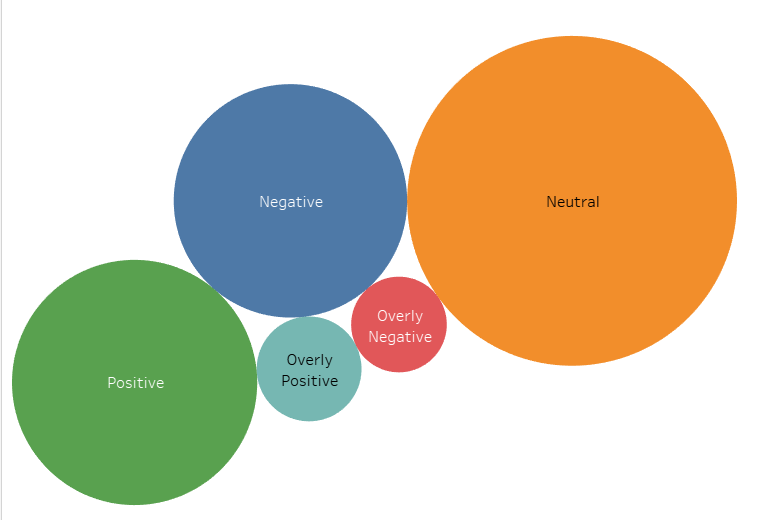


Figure 5: Sentiment graph

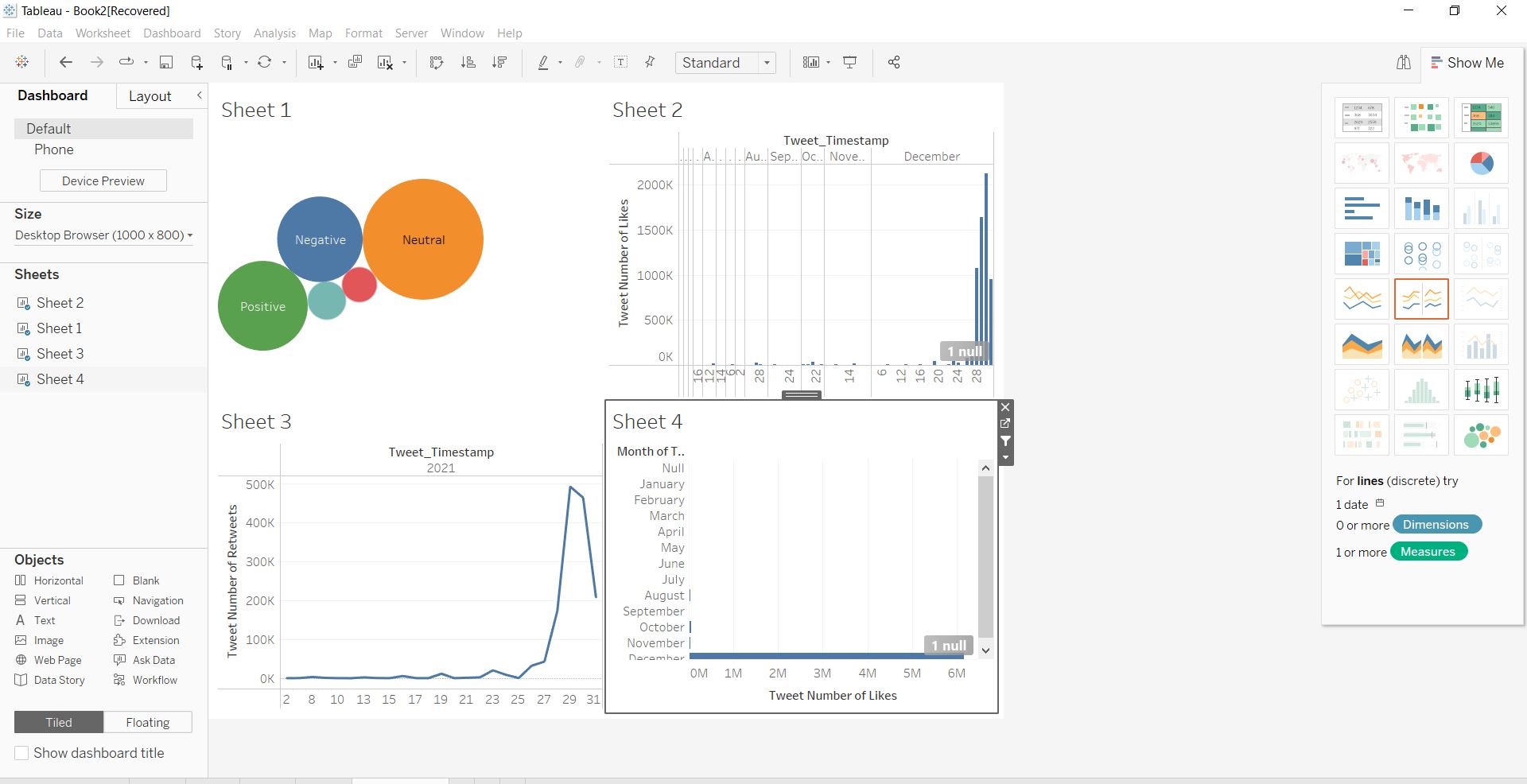


Figure 6: relation with sentiment and tweet

**6. Conclusion**

Pfizer is a well-known brand. The brand that produces the COVID-19 vaccines has been the subject of constant social media discussion for the past 12 months. Pfizer is often cited by reporters, politicians, and regular people on Facebook, Twitter, and TikTok as the epidemic continues to make headline news. The company didn't fully come to the attention of the public until around May 2020. At that time, the business disclosed that it was developing a COVID vaccine. When the business acquired FDA Emergency Use Authorization for their vaccine in December 2020, Pfizer reached its peak. The first vaccination was given a few days later. Pfizer-related posts filled social media. With 172 years of expertise, a substantial marketing budget, and the means to reach millions of people, Pfizer is well-equipped. Results show that the opinion of Pfizer/BioNTech remained favourable when compared to the emotion toward the other COVID-19 vaccines. Periods or spikes of unfavourable sentiment were also experienced by the Pfizer/BioNTech vaccine.

**7. Limitation**

It's vital to keep in mind that Twitter does not accurately represent the world's population, despite the fact that our study respected the sources and data in great detail. There are a large number of people who do not currently utilize Twitter. An overwhelming majority of Twitter users—44% of all users—are in their 30s to 49s. (Barnhart, 2021) Twitter data does not give a broad overview of the topic. Slang, typos, and other unclear terms in the data set cannot be understood by the sentiment function that was implemented in Python (Rahman et al, 2020). Additionally, the study used a relatively small sample of 2022 tweets from around the world.

Using free data scraping tools, it becomes very difficult to generate necessary data for this paper to tell a business story. Hardware also is not enough, no scraping data.

8. References

1. Prasanna (2020). Explain the use of Selenium Webdriver with Python. [online] Medium. Available at: https://prasannajyothi.medium.com/explain-the-use-of-selenium-webdriver-with-python-d7fc8e2d66d3 [Accessed 8 Dec. 2022].
2. Hwang, J.P. (2020). What Python package is best for getting data from Twitter? Comparing Tweepy and Twint. [online] Medium. Available at: <https://towardsdatascience.com/what-python-package-is-best-for-getting-data-from-twitter-comparing-tweepy-and-twint-f481005eccc9>.
3. Python, R. (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/>.
4. <https://www.trustradius.com/products/octoparse/>
5. Octoparse.com. (2019). Octoparse 3rd Anniversary！. [online] Available at: https://www.octoparse.com/.
6. www.octoparse.com. (n.d.). Get to Know Your Users Through Data Analysis. [online] Available at: https://www.octoparse.com/blog/get-to-know-your-users-through-data-analysis [Accessed 8 Dec. 2022].
7. **Nitesh V. Chawla, Kevin W. Bowyer, Lawrence O. Hall, W. Philip Kegelmeyer.”** SMOTE: synthetic minority over-sampling technique”, ©2002 AI Access Foundation and Morgan Kaufmann Publishers, published 06/02, 321—357.
8. Mee, A., Homapour, E., Chiclana, F. and Engel, O. (2021). Sentiment analysis using TF–IDF weighting of UK MPs’ tweets on Brexit. Knowledge-Based Systems, 228, p.107238. doi:10.1016/j.knosys.2021.107238.
9. Conner, C., Samuel, J., Garvey, M., Samuel, Y., and Kretinin, A., (2020) Conceptual Frameworks for Big-Data Visualization:Discussion on Models, Methods and Artificial Intelligence for Graphical Representations of Data. Handbook ofResearch for Big Data: Concepts and Techniques, Apple Academic Press, USA. Website:http://www.appleacademicpress.com/handbook-of-research-for-big-data-concepts-and-techniques/97817718898
10. (PDF) Social Media Sentiments Towards Vaccines. Available from: <https://www.researchgate.net/publication/353531634_Social_Media_Sentiments_Towards_Vaccines> [accessed Dec 02 2022].
11. www.pfizer.com. (n.d.). Pfizer and BioNTech Announce Data Demonstrating High Immune Response Following a Booster Dose of their COVID-19 Vaccine in Children 5 Through 11 Years of Age | Pfizer. [online] Available at: <https://www.pfizer.com/news/press-release/press-release-detail/pfizer-and-biontech-announce-data-demonstrating-high-immune>.
12. Samuel, J., Ali, G. G., Rahman, M., Esawi, E., & Samuel, Y. (2020). Covid-19 public sentiment insights and machine learningfor tweets classification. Information, 11(6), 31 (PDF) Social Media Sentiments Towards Vaccines. Available from: <https://www.researchgate.net/publication/353531634_Social_Media_Sentiments_Towards_Vaccines> [accessed Dec 08 2022].