

# Twitter Market Analysis of Health Supplements

---

Visualization of Wholepreneur lifestyle Keywords

Benjamin Esene (100350791)  
Soniya Gosavi (100343274)  
Andre Kozlowski Henrique (100350427)  
Tuan Anh Nguyen (100348136)  
Eduardo Solano-Salgado (100351079)  
Puja Unni (100352589)

CPSC 4820 | Visualization for Data Analytics  
Langara College

December 15, 2020

# Table of Contents

<b>Problem Definition</b>	<b>3</b>
Project Overview	3
Data Pipeline	3
Wholepreneur Lifestyle	3
About Wholepreneur Lifestyle	3
Social Media Analysis	4
<b>Results and Analysis</b>	<b>5</b>
Vegan + Organic Collection	6
Meal + Replacement Collection	7
Health Collection	8
Stress + Relief Collection	9
Competitors Collection	10
Collections with Insufficient Data	11
Product Branding Analysis	11
Limitations and Recommendations	11
<b>References of Data</b>	<b>12</b>

# Problem Definition

## Project Overview

It is known that the launch of a new product can be a challenging task for established companies. When talking about startups and small companies, it is possible to add an extra layer of difficulties since most of them lack the knowledge about the market they are entering. Thus, in order to decrease the chances of failing or making the wrong decisions, a market analysis is always an excellent first step.

Our project aims to support Wholepreneur Lifestyle, an organic/healthy product company, to acquire a better understanding of the market, by drawing insights through the research and sentiment analysis of social network posts (namely Twitter), using specific keywords related to their products and targeted market.

## Data Pipeline

During this first stage of our project, we decided to limit our research field to social networks, with an emphasis on the Twitter platform. In order to access social media information, we developed a python script to define the search parameters. This program is connected to the Twitter database through a free application programming interface (API) account. All information fetched from this platform was stored in a cloud-based MongoDB database. From this platform, we were able to access the information in order to proceed with further analysis (sentiment, location, timing, etc.). As our final step, we used two different software (Tableau and Microsoft Excel) to provide various visualizations of the data.

## Wholepreneur Lifestyle

### About Wholepreneur Lifestyle

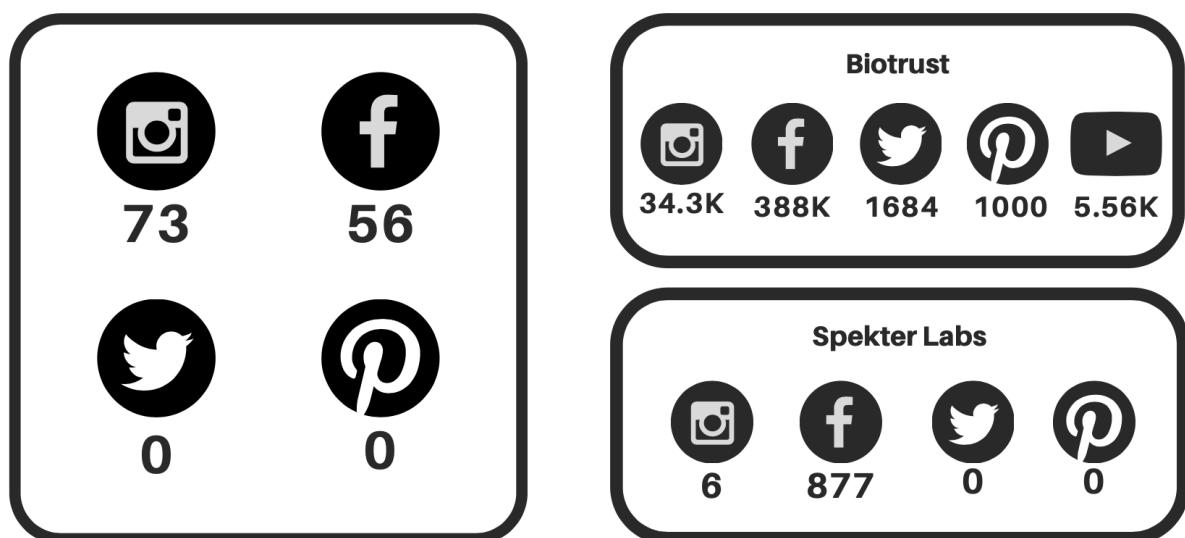
Entrepreneurs are major contributors to economic growth by developing products, services and innovative technologies. More so, the competition between entrepreneurs challenges existing companies to become more competitive which improves the overall productivity of a nation and provide new job opportunities. However, the productivity of entrepreneurs which is essential, given the high level of competition, depends on a sound mind to process information, a strong immune system to guarantee no days off due to illness, being able to perform at optimal level, being able to get enough rest to function effectively, and having a balanced overall health. Hence, the need to guarantee maximum level of productivity led to the birth of **wholepreneur lifestyle** for entrepreneurs by entrepreneurs who understand the daily struggles and challenges.

**Wholepreneur lifestyle** aims to transform the lifestyle of entrepreneurs by changing their habits and developing behaviours that align with high levels of productivity by using their certified organic and health products. Currently, they have 8 certified organic products known as total wellness supplements in the market with 2 products in the pre-launch phase

that target 5 key areas of focus designed to manage the critical needs of entrepreneurs: brain support, immune support, total body support, relaxation and protein/meal replacement.

## Social Media Analysis

We compared Wholepreneur lifestyle's social media presence with their competitors: Biotrust and Spekter Labs. According to the Figure 1 below, Biotrust has more than 34K Instagram followers, 338K Facebook followers, 1684 Twitter followers, 1000 Pinterest followers, and more than 5.5K Youtube subscribers. Moreover, Spekter Labs has only 6 Instagram followers and 877 Facebook followers. Conversely, Wholepreneur lifestyle has only 73 Instagram followers and 56 Facebook followers, a small number compared to the biggest competitor Biotrust.



**Figure 1. Social Media presence for Wholepreneur Lifestyle (Left), Biotrust and Spekter Labs (Right)**

Figure 2 below shows the engagement power from Wholepreneur's social media. We took a look at the latest 10 posts from both Facebook and Instagram social platforms. As we can see, the latest post from the company's Facebook page was on August 31, while the latest post from their Instagram page was on December 10 (Data obtained on December 14, 2020). According to the line chart, we can see the number of likes per post from the Facebook page were quite low as the maximum number of likes from the latest 10 posts was only 2 likes and there were 5 posts that had 0 likes. Therefore, the company's Facebook platform was not popular on social media. However, the company's Instagram platform performed better than their Facebook platform. The first 6 posts from the latest 10 posts have an average of 10 likes per post; however, for the last 4 posts, the trend went down with an average of 5 likes per post. Therefore, we can conclude that the followers did not engage much on Wholepreneur Lifestyle's social media platforms.

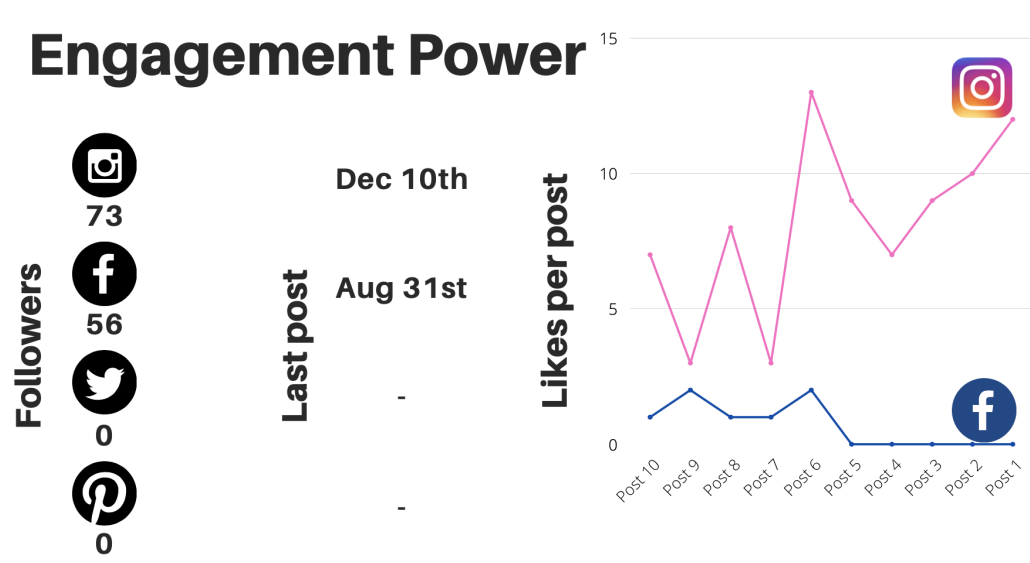


Figure 2. Wholepreneur Lifestyle Social Media Engagement

## Results and Analysis

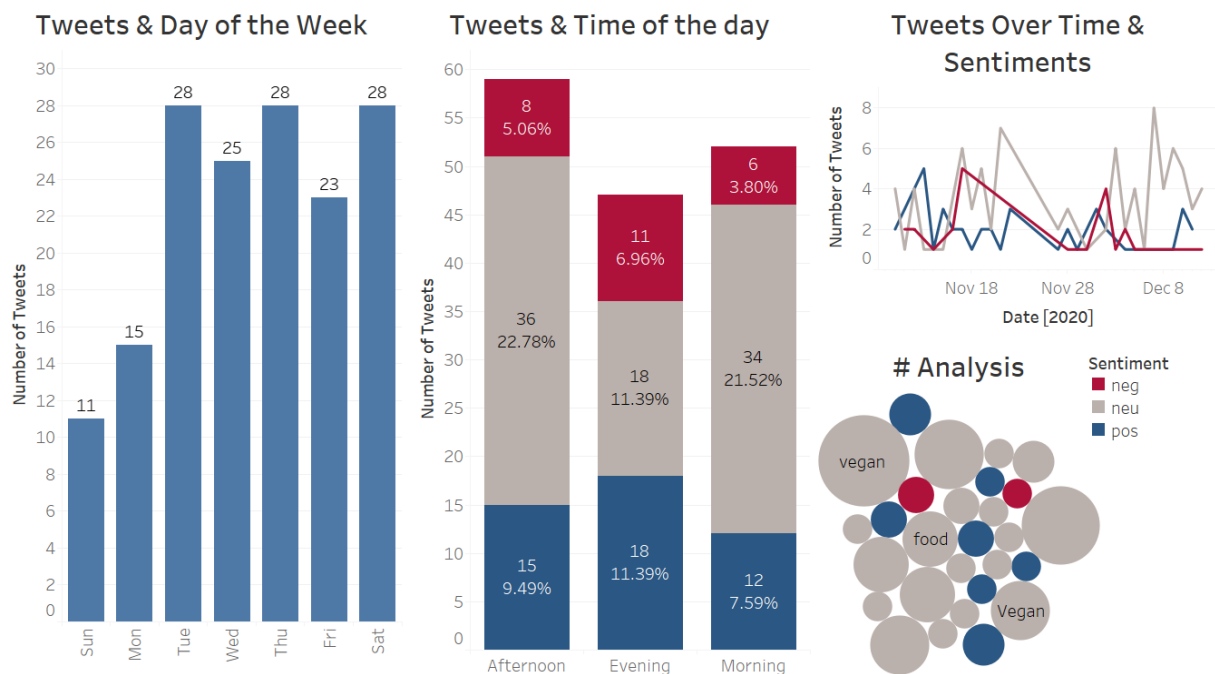
To better streamline our analysis, we created 8 different collections of Tweets, based on the commonality of the keywords used in our search parameters. Table 1 below shows the different collections used, their keywords, and total number of tweets obtained.

Table 1. Twitter Collections

Collection	Keywords	Tweets Count
Organic + Vegan	organic supplements vegan supplements	158
Meal + Replacement	meal replacement powder meal replacement protein powder organic meal replacement protein meal replacement protein powder meal replacement	111
Stress + Relief	mental calm supplement relaxation supplements stress relief supplements supplements for stress supplements for stress and fatigue supplements to reduce stress	15
Competitors	BioTrust SpekterLabs	38
Health	health and wellness products online immune support supplements immune system supplements supplement for endurance supplements for immune system	70
Brain Enhancement	brain enhancement supplements nootropics for creativity supplements for focus and concentration	2
Workout	all natural pre workout healthiest pre workout natural pre workout natural workout supplements	13
Others	supplements for work nootropics for sale	1

## Vegan + Organic Collection

The first collection in our analysis intended to analyze the growing vegan, plant based market, as well as public opinion of organic products. Figure 3 below shows the dashboard we obtained by analyzing the tweets from the vegan+organic collection.



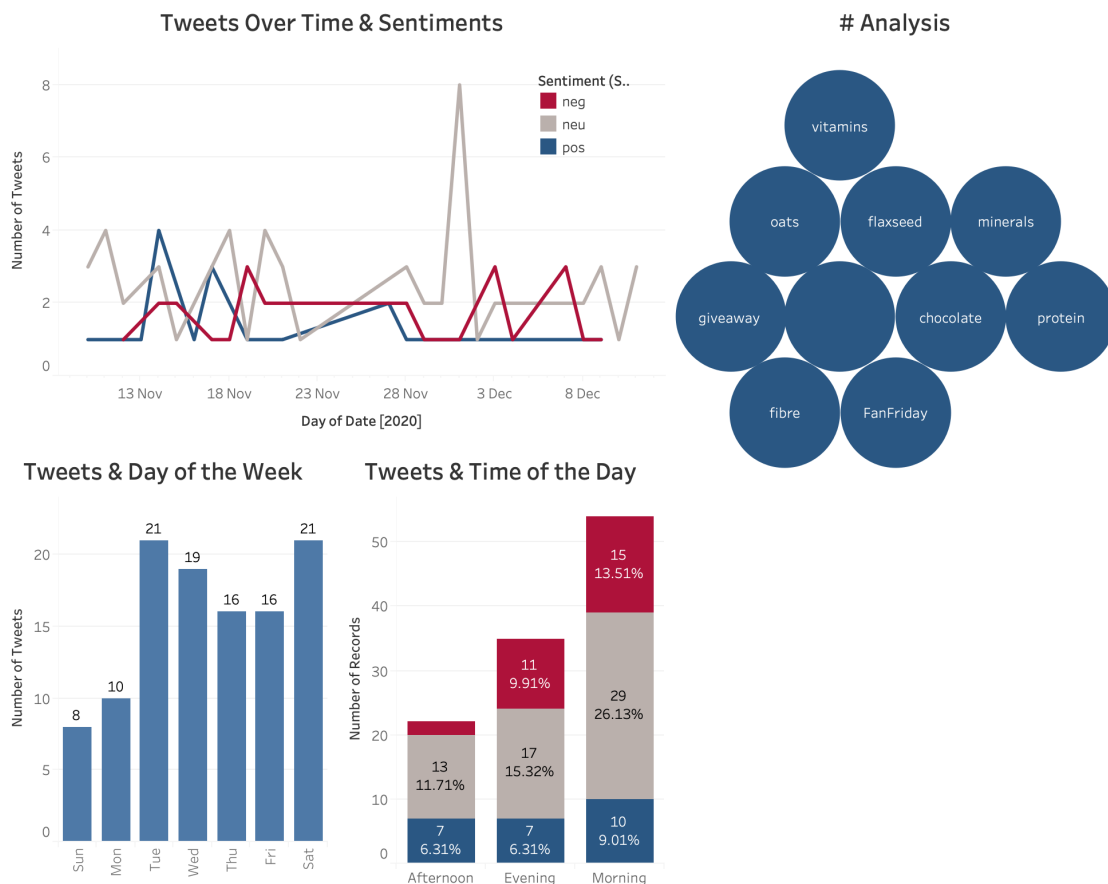
**Figure 3. Vegan + Organic Collection Dashboard**

From the Tweets & Day of the Week barplot we can observe a high twitter activity on Tuesday, Thursday, and Saturday, with a low count of posts on Sunday. The Tweets & Time of the Day stacked barplot helped us analyze the time of the day with the most twitter activity, as well as the distribution of positive, neutral, and negative tweets. We can see from the plot that the time with the highest overall activity is the morning (0 to 12 pm), while the largest percentage of positive tweets occurred in the evening hours (6pm to 12am). From this plot we can also infer the overall positive sentiment of tweets was 28%. The Tweets Over Time & Sentiments line graph shows the total number of tweets per sentiment throughout our one-month analysis. We can see that Twitter activity consisted of peaks and troughs of posts, with the highest overall activity in the days leading to the Black Friday weekend, and the days following. Lastly, we performed an analysis of the hashtags included in each of the Tweets in the collection. The bubble chart above (bottom right corner) shows the distribution of sentiment and number of times a hashtag was used. We can see that the majority of tweets were of a neutral sentiment, with the #vegan being the most used in the collection (note how the hashtag search is case sensitive).

Overall, we observed a trend of neutral tweets, likely due to the high activity of promotion/sales tweets leading to the Black Friday/Cyber Monday weekend. Additionally, contrary to what many may believe, Sunday did not show a significant level of activity, with most people posting during the week, which could be of interest to the company when deciding to launch a marketing campaign.

## Meal + Replacement Collection

This collection aimed at studying one of the most distinct product lines of the company: meal replacement products. Figure 4 below shows the dashboard of the collection.



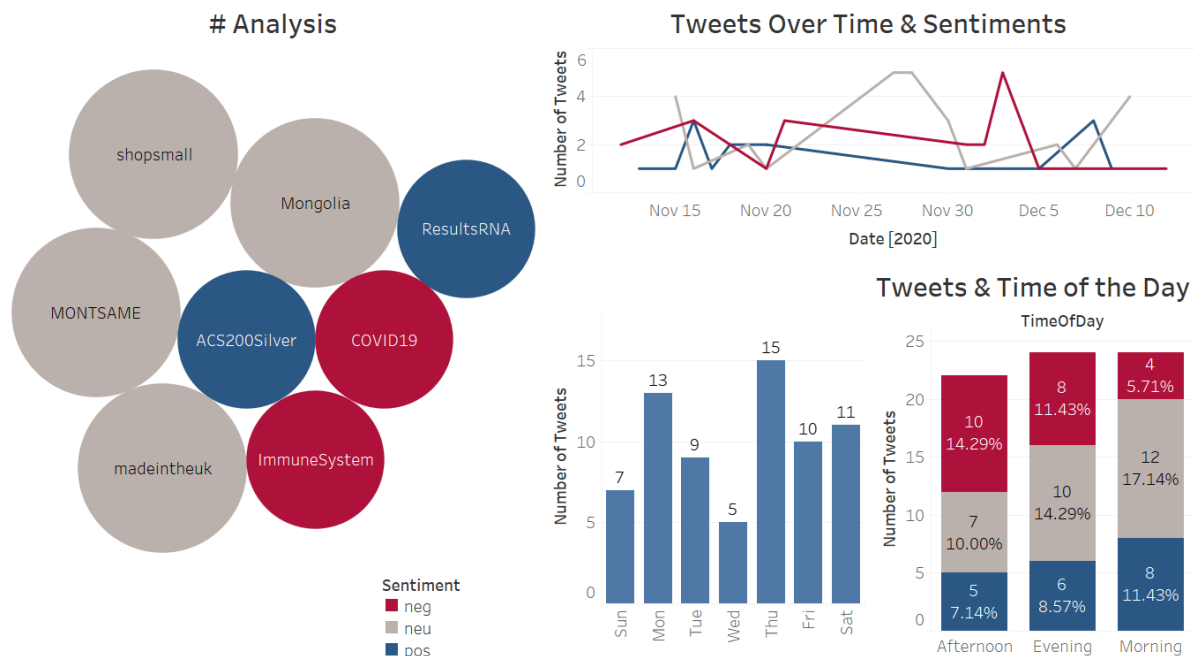
**Figure 4. Meal + Replacement Collection Dashboard**

From the Tweets & Day of the Week barplot we can observe a high twitter activity on Tuesday and Saturday, and a low count of posts on Sunday. From the Tweets & Time of the Day stacked barplot, we can see that the time with the highest overall activity and with the highest proportion of positive tweets is the morning (0 to 12 pm). From this plot we can also infer the overall positive sentiment of tweets was 22%. The Tweets Over Time & Sentiments line graph shows the total number of tweets per sentiment throughout our one-month analysis. We can see that there is a clear peak of activity around December 1. The hashtag analysis bubble chart shows the distribution of sentiment and number of times a hashtag was used. We can see that the majority of tweets were of a positive sentiment.

In general, there is a similar trend of low activity on Sunday as observed in the previous collection, and an overrepresentation of neutral tweets, likely caused by the large number of promotional tweets during the shopping season in late November. Also, there is not one particular hashtag that is used more often than others in the tweets we analyzed.

## Health Collection

The next collection we analyzed, included keywords related to the health and immune system-boosting supplements Wholepreneurs lifestyle offers. Figure 5 below shows the resulting graphical analyses of the collection.



**Figure 5. Health Collection Dashboard**

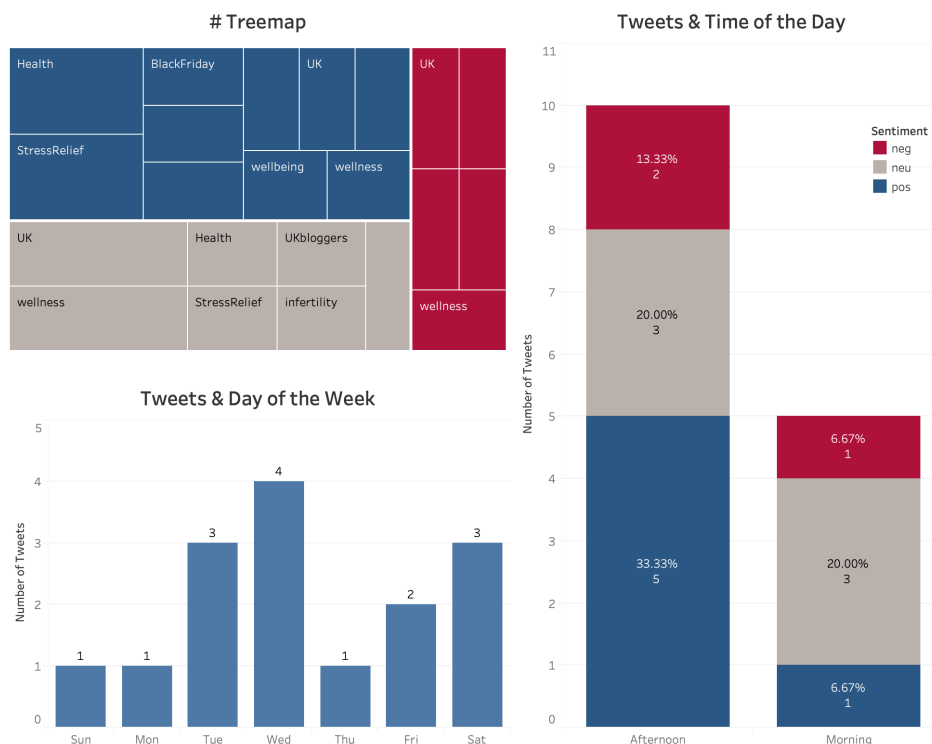
The Tweets & Day of the week barplot shows high activity on Thursday, and low activity on Wednesday and Sunday. The Tweets & Time of the Day stacked bar plot shows that the overall twitter activity for the collection was very similar throughout the day, with a slightly lesser amount in the afternoon hours (12pm to 6pm). The largest proportion of positive tweets occurred during the morning hours, with 11.4% of tweets assigned a positive sentiment. We can also infer that the overall percentage of positive sentiment for the collection is 27%. The Tweets over Time & Sentiments line graph shows a constant twitter activity throughout the month, with a clear peak of neutral tweets during the Black Friday weekend, and a peak of negative tweets around December 1. The hashtag bubble chart shows an even distribution of sentiments, with a slightly larger proportion of neutral tweets. Note how #COVID19 is present in the collection and its associated with negative tweets. This is likely due to the nature of the keywords used to gather tweets for this collection. Since we included keywords related to health and immune systems, it is likely that a few tweets regarding the COVID-19 pandemic made it to our collection. This hashtag analysis made it possible to realize this observation.

Overall, the trend of high activity during the week and Saturday is maintained in this collection, as well as the larger proportion of neutral tweets, likely due to the Black Friday weekend.



## Stress + Relief Collection

This collection aimed at analysing the most relevant keywords during this pandemic which includes supplements for stress, mental health, relaxation and fatigue. Figure 6 below shows the dashboard created for Stress Relief collection.

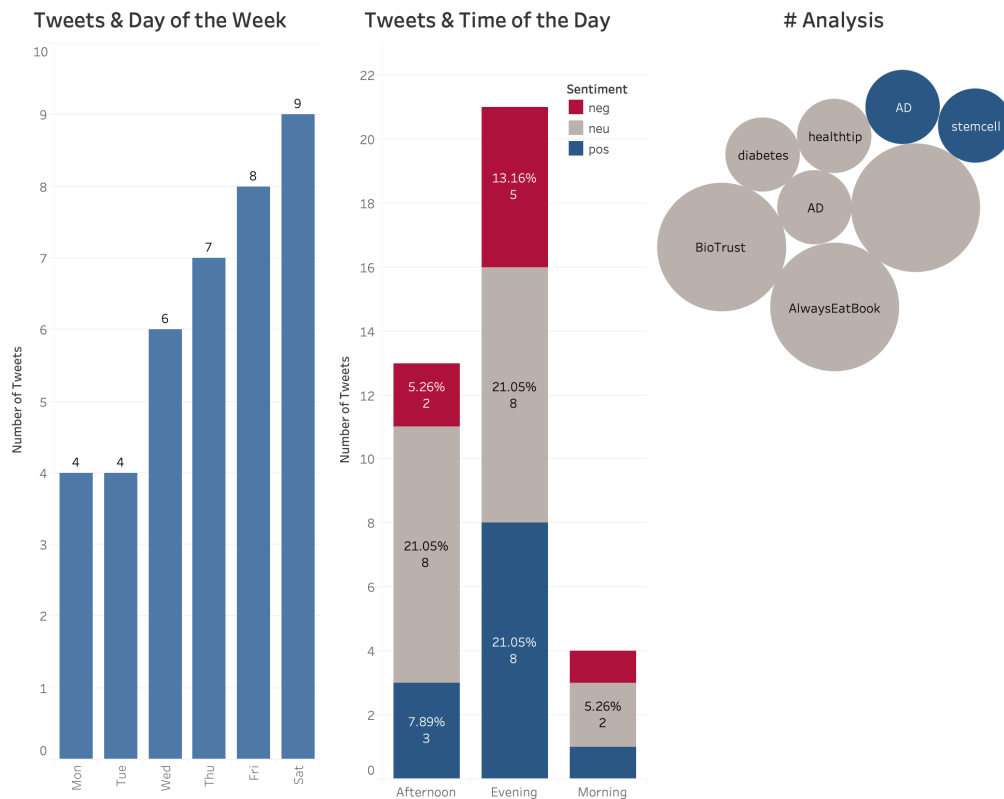


**Figure 6. Stress Relief Collection Dashboard**

The Tweets & Day of the Week bar graph shows high activity on Wednesday and low activity on Sunday, Monday and Thursday which makes sense because it could get stressful at work during mid-week and people may search the web for options to reduce stress and fatigue during this time. The Tweets & Time of the Day stacked bar graph indicates that the overall twitter activity for the collection was higher during the afternoon (12pm to 6pm) and there was no activity during the evening (after 6pm). The largest proportion of positive tweets occurred during the afternoon hours with 33.3% positive sentiment. We also observe that the overall percentage of positive sentiment for this collection is 40%. The hashtag treemap shows a similar distribution for positive and neutral sentiments and lesser proportion for negative sentiments. Note how #StressRelief appears in positive and neutral sentiment but the higher proportion in positive sentiment indicates that people spoke more positively about stress relief which is a good mindset to have in this situation. We can see that #UK appears in all three sentiments with the largest proportion of tweets being neutral which could be because there were a lot of promotional offers to sell products that reduce stress and balance mental health in the UK during the Black Friday week. Overall, the trend of activity is higher during the afternoon hours and during the mid week for this collection.

## Competitors Collection

Finally, we wanted to analyze public opinion and sentiment towards the 2 biggest competitors for Wholepreneur lifestyle: Biotrust and Spekter Labs. Figure 7 below shows the results of our Competitors collection analyses.



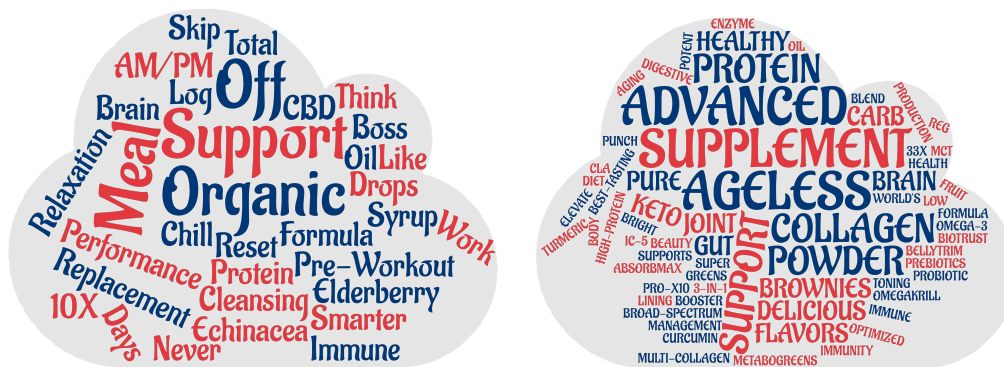
**Figure 7. Competitors Collection Dashboard**

The Tweets & Day of the Week barplot shows an increasing trend in activity as the week moves along, with 0 tweets on Sunday. The Tweets & Time of the Day stacked bar plot shows a high overall level of activity and the highest positive sentiment during the evening hours, and little activity during the morning. We can also infer from this plot that the overall positive sentiment of the collection was 32%. A hashtag bubble chart shows a greater proportion of neutral tweets to positive tweets, with the #BioTrust as one of the more significant hashtags in the collection. This is in part due to the fact that Biotrust has a company twitter page and it heavily marketed its products during the Black Friday/Cyber Monday shopping weekend.

Overall, we observed the same trend of high social media activity during the week and no activity on Sunday. Similarly, we observed a large distribution of neutral tweets in the collection, likely linked to sales and/or promotional tweets. Moreover, we decided not to include the line graph of tweets over time because the graph showed a flat line of tweets, and a single large neutral peak around the Black Friday weekend.

The last three collections we analyzed were the brain enhancement collection, the workout collection, and the others collection. As seen from Table 1, these collections did not fetch enough tweets to perform a proper analysis, during the one-month data-gathering time. This is partly due to the keywords used for these collections, or because these specific markets might be better analyzed using other social media platforms (the fitness and workout community has a significant instagram presence), where the results can be of more use to Wholepreneurs lifestyle to perform marketing analyses.

One of the key factors in marketing is to create a clear brand identity, especially when a company is entering a market for the first time. Having a clear product identity and offering an alternative to competitors offers a strategic advantage to a company like Wholepreneur lifestyle. The final analysis we performed focused on how well defined the product line is compared to the competitors. Figure 8 below shows the results of our analysis.



The word clouds above were created using a list of the official products from Wholepreneur lifestyle and its competitors (Biotrust and Spekter Labs). From the word clouds we can see that wholepreneurs lifestyle has a distinct line of products that show no significant overlap with the product line of its competitors. We see that there is a clear emphasis on organic products, as this is one of the major selling points of the product line of the company. On the other hand, we see that there is no mention of the word “supplements” in the actual product names, which could pose a challenge in product recognition, as it might be difficult for potential buyers to know the products are supplements, or it might affect the extent in which the product line appears after a quick Google search (“Vegan supplements” for instance).

The biggest limitation faced during our project was the free Twitter API used to obtain the data to be analyzed. The free API has a limit on the number of tweets that can be fetched per query, the maximum time the data can be obtained from, and the number of search parameters and information we can obtain per tweet. As such, our analysis was reduced to individual tweets, date of posting, hashtags, and retweets/favourites. The latter were not

included in our analyses, as these are boolean values (True/False) and offered little inference to the conclusions we wanted to draw. Another limitation we faced was the fact that we only included Twitter in our analysis. By focusing on one social media platform, there is potential for our analysis to leave out a significant proportion of the users the company wants to reach. One example of this is the fitness/workout market, which showed almost no twitter activity, but there is a known significant instagram activity. The final limitation we faced was in the sentiment analysis algorithm used to assign emotions to the tweets in the collections. Our sentiment analysis algorithm initially assigned only positive and negative values to the messages, and we noticed that many of the sales/promotion tweets were assigned negative values, resulting in a skewed distribution of sentiments.

In order to improve our analysis we recommend to include other social media platforms to get a more well-rounded view of the market potential of the product line. Additionally, having an official twitter page would allow us to analyze customer engagement and company-related sentiment directly, as our current analysis only used general sentiment towards keywords associated with products. Having a stronger social media presence in the platform of study would allow for an analysis of the brand, as well as brand engagement. Finally, when looking at the official instagram page, we noticed that the posts that generated the most community engagement showed images of people enjoying the wholepreneur lifestyle, instead of a direct focus on the products. We believe that marketing the lifestyle associated with the brand instead of only the product could lead to higher consumer engagement.

## References of Data

All individual tweet messages used in the present report were obtained using the official Twitter API and compiled in JSON files for analysis. Information regarding the activity of social media engagement for Wholepreneur lifestyle and competitors was obtained from their official social media pages. General Wholepreneur lifestyle information was obtained from the official company website. Keywords and relevant industry data was initially provided by the client and instructor (excel, word file). Word clouds were done using the free online tool at wordclouds.com. Dashboards were created using the Tableau® Desktop software. No sensitive or proprietary information was released or exchanged during the time of this project.