A logo for a college

Description automatically generated

**Vikram Jeet Singh**

**Capstone Project Report**

**Health Analytics Student**

**A00290100**

**Table of Contents**

[**Introduction** 3](#_Toc169733435)

[**Scope of the project** 3](#_Toc169733436)

[**Objective of the project** 4](#_Toc169733437)

[**Specific technical project requirements** 4](#_Toc169733438)

[**Deliverables and milestones set with the client** 4](#_Toc169733439)

[**The outcomes and the satisfaction of the client** 5](#_Toc169733440)

[**The weekly challenges, the relationship with the client, and the feedback received** 6](#_Toc169733441)

[**Lessons learned, what experience did I take from this** 7](#_Toc169733442)

[**What I have done differently, and what was well done** 7](#_Toc169733443)

[**My contributions to the teamwork, what I have accomplished** 7](#_Toc169733444)

[**Results for the client, is the project a success or a failure** 8](#_Toc169733445)

[**Recommendations and conclusion** 8](#_Toc169733446)

[**References** 8](#_Toc169733447)

[**Appendix** 8](#_Toc169733448)

# **Introduction**

## **Project and client’s Name**

Coffee fruit (synonym: coffee cherry and coffee berry) is a fruit that covers coffee beans and has beneficial activities including antioxidant, anti-inflammatory, neuro-protective, energy Boost, maintains blood pressure and weight loss, and anti-aging properties. It is edible and has a short half-life. The main active constituents present in the fruit are Chlorogenic acid, Caffeine, and other antioxidants.

Cofilife (formerly known as CaffeBerry) provides a range of products showcasing the versatile nature of Coffee fruit (mainly focusing on Chlorogenic acid) possessing as a health supplement and promoting health and sustainability in one frame. The goal of the firm is to promote wellness, quality, and sustainability. The firm is highly innovative and offers various products, including Natural Health supplements or capsules, Protein Powders, Energy Bars, and Alcoholic/Non-alcoholic drinks. In general, these products are suitable for the age population between 18-40 years, but the positive effects are also seen in the older age population (Cofilife, 2024).

As a part of the analytics capstone project, Cofilife (especially, Mr. Shabs and Ms. Frida, who are the point of contact and clients) and my Institute (Cambrian College) provided me with an opportunity to work under them. In this report, I have briefly discussed the project, analysis, and results and the proposed keywords that can be used for the endorsement of product.

# **Scope of the project**

The scope of the project was to fetch and analyze the data using different social media (i.e., YouTube, Twitter, Pinterest, and so on), and use this data to identify the relevant hashtags for the product endorsement. It aims to promote coffee-fruit-based products that promote health and sustainability. The project includes obtaining data from social media using Python, R studio, or MS Excel for market research, promotional activities, and community engagement.

# **Objective of the project**

The objective of the project is to identify the relevant hashtags that are beneficial and aligned with the brand and product. Furthermore, hashtags which helps to enhance product awareness and reach among the target population, thus increasing the reach of CofiLife's products. Additionally, the project sheds light on health benefits and sustainability and opens the door for further market research.

**Limitations or constraints**

The major limitation of this project is data restriction and platform algorithm. Many platforms, including Twitter, LinkedIn, and some others, are highly confined and have a lot of data privacy restrictions, which restricts the analyst from fetching the data for analysis. As far as this report concerns, the extraction of specific data from social media, such as YouTube, being used in this project is of high quality, but was less available due to the novelty of the Cofilife products.

# **Specific technical project requirements**

The data analysis tool Python has been used for data extraction and analysis. The YouTube API key has been used to extract data from social media. Various libraries, such as pandas and IPython.display, are used for data manipulation and visualization, and convert json to readable format. For Visualization, matplotlib.pyplot and matplotlib.ticker is used, and for word cloud visualization and natural language processing, word cloud, nltk, nltk, corpus, and nltk, Tokenize has been used. Furthermore, seaborn is used for statistical data visualization.

# **Deliverables and milestones set with the client**

For the first three weeks, me and my group had tried and successfully identified the social media and fetch the API key from the relevant social media platforms. For the next two weeks (i.e., weeks four and fifth), our group tried to fetch the data week; for the sixth week, my group analyzed the data, and in the last week, the report was submitted to the client and institute. The milestones or deliverables are provided in the form of a final document and Python code and are to be submitted by June 20, 2024. The final document contains the brief and recommended list of hashtags and analysis. The Python code includes a detailed list of hashtags, their association with likes, views, and comment counts, hashtags that are being used in the top 50 videos, and Word Cloud stating the most frequent hashtags used for product promotion. This helps to increase the reach of the targeted customers and enhance social media engagement.

# **The outcomes and the satisfaction of the client**

As per the Client's guidance and considering the area of interest, python code was prepared, data was fetched, and analysis was generated. This section will briefly discuss the fetched data, analysis, and potential keywords and hashtags that the Client can use for the relevant purpose.

Firstly, the data was generated using the keyword "health and coffee," and the following data was generated (See Appendix Table 1 and Table 2). Tables 1 and 2 show the top 10 channels, related titles, and hashtags by views and a summary of the top 10 channels with total views, video count, and total tags used.

The table generated provides a good understanding of that what hashtags or keywords can be used or potentially be helpful to develop product image. Few of the examples are Few of the examples are 'Coffee', 'Health', 'Research', 'Science', 'Wellness Wednesday', 'Weight Loss Freak', 'health and wellness', 'healthy diets', 'weight loss tips', 'weight loss', 'health tips', 'wellness', 'natural remedies', 'healthy living', 'lifestyle tips', 'why you should drink black tea every day', 'black tea health benefits', 'black tea for weight loss', 'black tea heart health', 'black tea gut health', 'black tea stress relief', 'black tea blood sugar', 'health perks of black tea', 'daily black tea benefits', 'weight loss freak black tea', 'healthy eating', 'weight loss exercises at home', '9 unexpected benefits of coffee', 'benefits of coffee', 'coffee', 'coffee benefits', 'coffee health benefits', 'coffee uses', 'facts about coffee', 'health benefits of coffee', 'benefits of drinking coffee', 'coffee facts', 'coffee health facts', 'coffee health effects', 'is coffee healthy', 'is coffee bad for you', However the detailed keywords have been provide in the code (Please refer the appendix to access the code). Table 2 provides the channel title and total video count used in the analysis. However, It failed to provide nuance for how many tags to be used to get more views. For example, the channel title "Bright Side" contributed one video and 18 tags but has maximum views. On the other hand, the Channel titled "motivational doc" has four videos with 78 tags and is in the top three.

To get a clear perspective, a plot (please refer to appendix figure 1) is generated, and the relationship between the Total No. of tags and views, total no. of views and likes, and total no. of tags and comments, and the results show that if the number of tags is used between 30 and 40, views, likes and comments will proportionally increase. Furthermore, plot representing top-performing videos by view count have also been stated in the code (Please refer the appendix to access the code) and the following keywords may act as potential keywords for the product build-up and include 'health care', 'coffee', 'coffee benefits', 'facts about coffee', 'improve brainpower', 'lose weight', 'caffeine', 'blood pressure', 'immune system', 'get rid of stress', 'improve memory', 'coffee recipes', 'coffee shop', 'coffee beans', 'coffee history', 'truth about coffee'.

To get insight into popular and most used keywords and hashtags, the word cloud is generated (see appendix Figure 2, 3, and 4), and the result shows that keywords like diabetes, nervousness, lifestyle, coffee benefits, weight loss, stimulant, awake, fitness, pain, energy, healthcare, anxiety loss, wellness, body improve, addiction, positive, natural, osteoporosis, allergies, indigestion, wake, longevity, freak, engagement.

To conclude, the keywords mentioned in the analysis have the potential to gain the upper hand in the market, and the Client was satisfied with the results by going through the analysis.

# **The weekly challenges, the relationship with the client, and the feedback received**

The data collection challenge is the main challenge encountered during the project. Data extraction from LinkedIn and Twitter is highly challenging and is highly protected as per the privacy policies of the platforms. The second major challenge includes the extraction of relevant information that is aligned with the product and market needs.

The Cofilife or client is highly optimistic, and I appreciate the way he and my group have guided me throughout the project. The feedback, including switching from closed platforms to open platforms, such as YouTube, helped me to handle the project and challenges well. Other feedback, such as including summary analysis and generating analysis using more relevant videos instead of a playlist, is highly encouraging. Overall, feedback was highly positive and constructive and helped to deal with challenges like Data collection and technical issue during the project through a continuous communication channel.

# **Lessons learned, what experience did I take from this**

I have learned a different aspect of market research, particularly when it comes to brand and product buildup. One needs adaptability and flexibility to handle unexpected challenges and should be able to adapt as per the market requirement or demand; campaign success depends on community engagement and whether the product reaches potential customers. Of all the above, to get the upper hand in the market, high-quality and compelling data is essential to get the recognition in the market.

The key takeaways, such as understanding the market behavior using data, are the secret of modern-era product positioning, and In general, continuous effective communication helps to deliver better outcomes.

# **What I have done differently, and what was well done**

I would have incorporated more precise keywords to get more accurate word clouds. The other minor change that could have been made in the code is making it more flexible so that it can take different keywords at the same time and generate the keywords or word clouds more accurately.

I have provided detailed insight and the hashtag, which could prove essential in brand building. Furthermore, analysis using a plot, finding the relationship between hashtags vs view count or, comment count or like count, could be proven as an essential in building the product image. Furthermore, word cloud provides a list of all words that are frequently used to promote similar products.

# **My contributions to the teamwork, what I have accomplished**

Data fetching, generation, and analysis using the multifactorial approach, strategy development, setting up meetings, timely communication with clients and institute, and report submission suggestions by providing brief insight into work were some contributions or tasks that I did during the project.

# **Results for the client, is the project a success or a failure**

I believe the project is a success and meets the client's requirements, and the analysis could prove to meet the market by reaching targeted customers.

# **Recommendations and conclusion**

I highly recommend the use of analysis to gain an advantage in the market. Furthermore, from a future perspective, other platforms and code flexibility should be focused on meeting the requirements.

            To sum up, the project effectively identifies the key hashtags, and by using these hashtags, Cofilife's product will reach the targeted customers, benefiting society by promoting sustainability and wellness.

# **References**

Cofilife (n.d.). About CofiLife. Retrieved June 17, 2024, from https://cofilife.com/.

# **Appendix**

|  |  |  |
| --- | --- | --- |
| Channel Title | Title | Tags |
| Click On Detroit | Local 4 | WDIV | How drinking coffee impacts your health | ['Coffee', 'Health', 'Research', 'Science', 'Wellness Wednesday'] |
| Mayo Clinic | Mayo Clinic Minute: Health Benefits of Coffee | ['Mayo Clinic (Hospital)', 'Health Care (Issue)', 'Healthcare Science (Field Of Study)', 'Mayo Clinic Minute', 'coffee', 'caffeine'] |
| Good Morning America | New study shows the health benefits of coffee | ['Ashton', 'Dr.', 'Jen', 'benefits', 'coffee', 'health', 'new', 'p\_cmsid=2494279', 'p\_vid=news-85086039', 'study'] |
| WXYZ-TV Detroit | Channel 7 | Healthiest Types of Coffee with Dr. Oz | ['Detroit', 'Local News', 'distributable'] |
| Doctor Mike | The Ugly Truth About Coffeeâ€™s Effects On Your Body | ['doctor mike', 'dr mike', 'drmike', 'dr. mike', 'mikhail varshavski', 'doctor mikhail varshavski', 'mike varshavski', 'coffee', 'starbucks', 'coffee facts', 'facts about coffee', 'anxiety', 'decaf', 'caffeine', 'caffination', 'drink coffee', 'drinking coffee', 'too much coffee', 'dehyradration', 'cold brew', 'tea', 'iced tea', 'espresso', 'latte', 'dunkin', 'dunkin donuts', 'starbucks coffee', 'dunkin coffee', 'blood pressure', 'high blood pressure', 'weight gain', 'weight loss', 'weight management', 'diet', 'coffee weight loss'] |
| Doctor Mike Hansen | Why Coffee Is Killing You...Slowly | ['Why Coffee Is Killing You', 'acrylamide in coffee', 'Is coffee good or bad for your health?', 'coffee dangerous', 'drinking coffee', 'coffee', 'coffee health', 'coffee energy', 'doctor mike', 'mike hansen', 'doctor mike hansen'] |
| ZOE | Is Coffee Healthy? | James Hoffmann and Professor Tim Spector | ['JIMSEVEN'] |
| Dr. Eric Berg DC | 8 Unexpected Benefits of COFFEE You've Never Heard Before | ['9 unexpected benefits of coffee', 'benefits of coffee', 'coffee', 'coffee benefits', 'coffee health benefits', 'coffee uses', 'facts about coffee', 'health benefits of coffee', 'benefits of drinking coffee', 'coffee facts', 'coffee health facts', 'coffee health effects', 'is coffee healthy', 'is coffee bad for you', 'drinking coffee', 'coffee benefits dr berg', 'coffee dr berg', '9 benefits of coffee', 'coffee effects', 'coffee side effects', 'black coffee benefits', 'dr. eric berg', 'eric berg', 'dr. berg', 'berg'] |
| Weight Loss Freak ! | WHY YOU SHOULD DRINK BLACK TEA EVERY DAY: 5 AMAZING HEALTH PERKS! ðŸŒ¿ðŸ’ª #shorts #blacktea #inspired | ['Weight Loss Freak', 'health and wellness', 'healthy diets', 'weight loss tips', 'weight loss', 'health tips', 'wellness', 'natural remedies', 'healthy living', 'lifestyle tips', 'why you should drink black tea every day', 'black tea health benefits', 'black tea for weight loss', 'black tea heart health', 'black tea gut health', 'black tea stress relief', 'black tea blood sugar', 'health perks of black tea', 'daily black tea benefits', 'weight loss freak black tea', 'healthy eating', 'weight loss exercises at home'] |
| WXYZ-TV Detroit | Channel 7 | Ask Dr. Nandi: Is decaf coffee harmful to health? | ['Detroit', 'metro Detroit', '7 Action News', 'Ask Dr. Nandi', 'Is decaf coffee harmful to health?', 'WXYZ', 'decaf coffee'] |

**Table 1. Represents top 10 channels, related title, and hashtags by views**

|  |  |  |  |
| --- | --- | --- | --- |
| Channel Title | Total Views | Video Count | Total Tags |
| BRIGHT SIDE | 13596855 | 1 | 18 |
| AsapSCIENCE | 9462414 | 2 | 88 |
| motivationaldoc | 7856956 | 4 | 78 |
| Doctor Mike | 4829479 | 1 | 35 |
| HealthNormal | 3030343 | 1 | 16 |
| Sean Nalewanyj Shorts | 2857533 | 1 | 14 |
| Sean Hashmi MD | 1941378 | 1 | 27 |
| The Infographics Show | 1561856 | 1 | 0 |
| Dr. Eric Berg DC | 1464948 | 2 | 50 |
| Dr. Josh Axe | 1257057 | 1 | 14 |

**Table 2. Represents top 10 channels, total views, and video count and total tags used**

A graph of data on a white background

Description automatically generated

**Figure 1. Plot representing relationship between Total No. of tags and views, Total No. of view and likes and Total No. of tags and comments**

A close up of words

Description automatically generated

**Figure 2. Figure representing various suggested keywords by views to be used for the endorsement of products.**

A close-up of words

Description automatically generated

**Figure 3. Figure representing various suggested keywords by using top 50 videos by views to be used for the endorsement of products.**

A close up of words

Description automatically generated

**Figure 4. Figure representing various suggested keywords, which are highly relevant to product by views, to be used for the endorsement of products.**