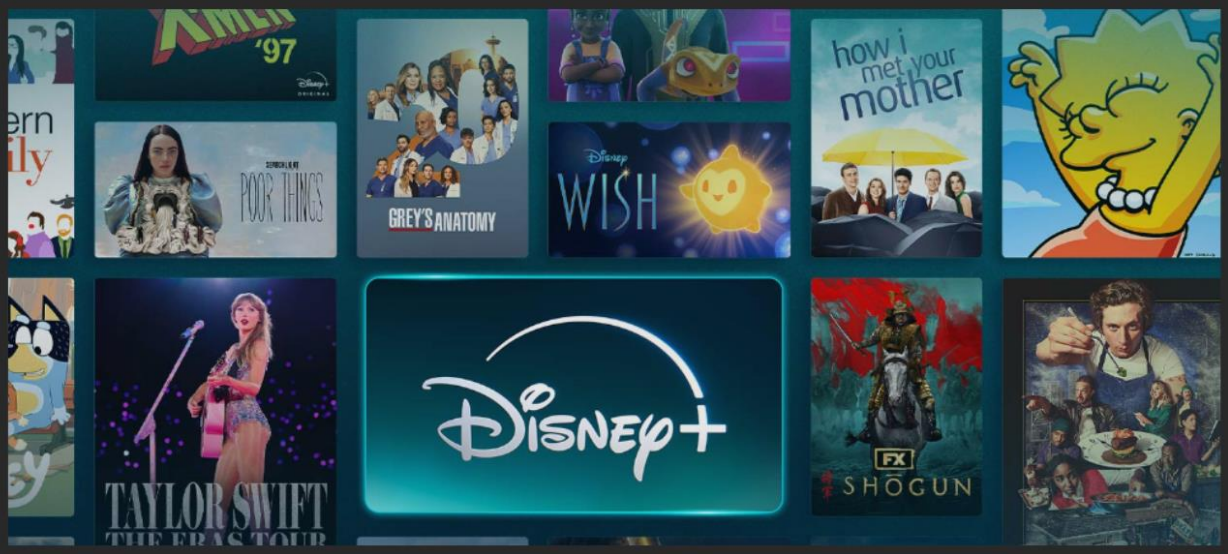

Student Id: 20549904

BUSI4370 UNUK SPR1 23-24

Analytics Specializations and Applications

Coursework 2 - Conversation and Network Analytics

University of Nottingham

Twitter Data and micro-influencer Analysis for DisneyPlus Brand:

1. Executive Summary:

A comprehensive analysis of Twitter data focusing on Disney Plus was conducted, aiming to uncover insights regarding brand sentiment, influencer engagement, and pertinent topics over a span of 10 days from 22nd February to 3rd March 2023. The sentiment analysis revealed a predominantly positive to neutral sentiment towards the brand, with only 10% of the online community being negative.

Topic modeling was performed using three different methods: KMeans & TF-IDF, NFM & TF-IDF, and LDA. Across all methodologies, three primary topics emerged from the online discussions: the Disney Plus series "The Mandalorian" (a highly popular series set in the Star Wars universe, scheduled for release on March 1, 2023), "Black Panther" and Marvel Studios, and discussions about various streaming platforms. A significant increase in tweet volume was observed on March 2, 2023, coinciding with the release of the third season of "The Mandalorian." Clustering analysis was conducted to discern patterns in online engagement and overall user behavior, resulting in the identification of three distinct segments.

Later, a micro-influencer for the brand was identified based on engagement metrics such as tweet count, followers, favorites, and retweets. This user positively impacted the brand, actively posting for 7 out of the 10 days analyzed. And also authored the most favorited and retweeted tweet during this period which centered around the globally popular KPOP group, BTS. The influencer consistently engaged with topics such as "The Mandalorian," Marvel, and other latest releases. On further analysis, it was found that it was the official Disney Plus twitter account. So, additional criteria based on influential tweets and other engagement factors was considered to identify the second choice for influencer.

2. Approach breakdown:

Tweets Cleaning: In the preprocessing phase, the data underwent several steps to enhance its quality. Stopwords were eliminated, tweets were converted to lowercase, and any redundant spaces, punctuation, or special characters were removed using regular expressions.

Location-wise analysis: A significant portion of the dataset, consisting of 972 records out of 2545, lacked location information. Additionally, 128 locations were identified using alphanumeric characters, further complicating the analysis. Despite these challenges, Los Angeles emerged as the most frequently mentioned location, appearing 147 times. To address this, a comprehensive list of states, major cities in the USA and South America, popular cities in the UK, and other significant countries such as India, Germany, Spain, and Canada were employed in a conditional loop to identify popular locations.

Sentiment Analysis: Utilizing the NLTK and TextBlob library, sentiment labels were assigned to each tweet. Subsequently, deeper sentiment analysis was conducted based on these labels, categorizing tweets as positive, negative, or neutral.

Topic Modelling: Three distinct methods—KMeans, NFM with TF-IDF, and LDA—were employed for topic modeling. The common terms such as 'Disney' and 'DisneyPlus' were designated as stopwords. In the case of KMeans, TfidfVectorizer was utilized with specific parameters, including max_df = 0.8 and min_df = 2 (ignoring all words with term document frequency more than 80% and appearing in less than 2 documents), while unigrams and bi-grams were identified to offer a more comprehensive insight into the discussed topics.

Twitter Author Clustering Analysis with KMeans: Utilizing PCA and KMeans, Twitter authors were clustered into three segments based on tweet count, follower count, retweets, and favorites. This segmentation revealed distinct user groups with varying online behavior and engagement patterns.

Micro-influencer analysis: Engagement rates for individual users were computed based on metrics such as tweet count, retweets, favourites, and follower count. These metrics were leveraged to identify key influencers, while an

influencer score, calculated as the sum of favourites and retweets received, was employed to gauge the level of engagement generated by specific tweets within the online community.

Engagement rate was calculated as: Engagement_rate = (no. of retweets + no. of favourites)/ no. of followers

3. Data Description:

The Disney plus Twitter dataset has 2545 records with 15 features. The data ranges for 10 days from 22nd February to 3rd March 2023. On counting the no. of tweets made each day, it was found that there is no data of tweets made on the 24th, 27th, and 28th of February 2023 or that no tweets related to Disneyplus were made during those days. So, the data for only 7 active days is available, refer to Fig. 1.

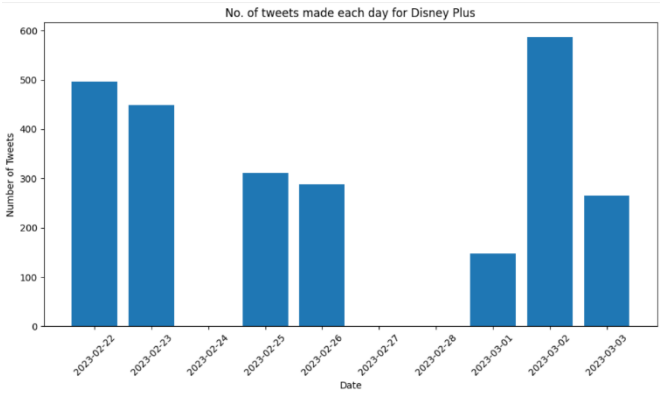

Fig 1

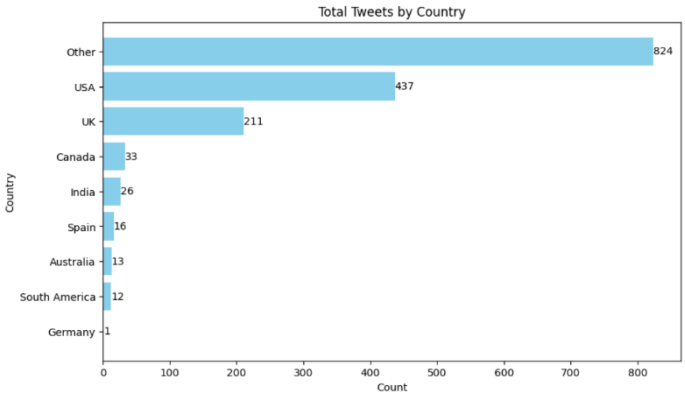

Fig 2

Key data transformations and observations include renaming the 'favourites_count' column to 'tweets_or_status_favorited_count' to avoid confusion with the 'favourite_count' column. Null values were only present in the 'location' column, with 972 missing values. Of the remaining locations, 437 tweets originated from the USA, 211 from the UK, while 824 were from unknown (miscellaneous) locations. Among the non-null locations, 'Los Angeles, CA, USA' emerged as the most popular, refer Fig 2. Additionally, 128 alphanumeric locations were identified.

The dataset consists of **1,073 unique users**, with **40 verified** and **1,033 unverified** users. Most verified users exhibited low activity, with only 4 out of 40 posting more than 10 tweets during the analysis timeframe. Conversely, the top 10 most active (posting most tweets) users were all unverified. The dataset recorded a substantial volume of tweets, peaking at 587 on March 2nd.

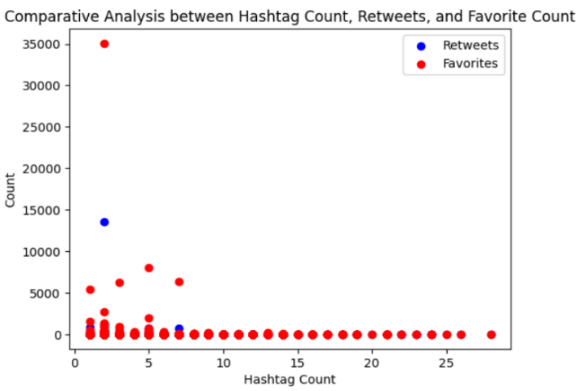

Fig 3

Among 2,922 unique hashtags in the tweets, **#DisneyPlus** was the most popular one. The other top hashtags included **#disneyplus**, **#TheMandalorian, #StarWars, #Disney, #Netflix** etc. However, an <u>inverse relationship between the number of hashtags used and retweets or favourites received was observed,</u> refer Fig 3. Usually, hashtags serve to amplify the reach and visibility of tweets. However, during this specific period, a contrary trend is observed. Surprisingly, even the most widely circulated tweet contained merely two hashtags. Notably, among all the tweets analyzed, only 184 directly promoted the brand through the use of brand tags, such as **@Disneyplus, @disneyplus, or @DisneyPlus**.

4. Brand Exploration:

In this section, Disney Plus's Twitter presence is explored, dissecting user engagement, sentiment, and trending topics. Data-driven methodologies are employed to uncover insights into brand perception and audience interaction.

A. Brand Sentiment analysis and topics discussed:

As discussed previously, sentiment analysis of tweets was conducted using the TextBlob library, revealing an overall sentiment for the brand that ranged between positive and neutral, refer Fig 4. Daily analysis of top tweets consistently indicated a positive tone across the board, with discussions primarily revolving around recent releases.

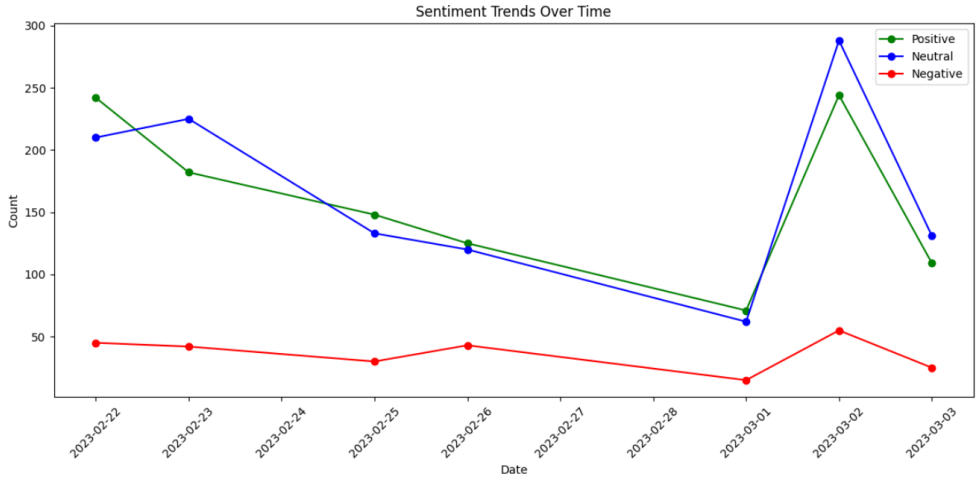

Fig 4

Upon comprehensive sentiment analysis of all tweets, it was observed that there are 1121 positive, 255 negative, and 1169 neutral tweets. Notably, positive discussions often centered on streaming platforms like Netflix, Hulu, Spotify, and the Mandalorian series. However, negative sentiments, primarily associated with Star Wars, may have been influenced by the release of episodes from the series "Star Wars: The Bad Batch" during the same timeframe, potentially leading to misidentification. It is important to note that tweets labelled as negative may include mentions of the "Bad Batch" series, where the term "bad" could be misinterpreted, thus, further investigation into the context of such tweets is needed. Meanwhile, neutral discussions frequently focused on topics related to "The Mandalorian" and the broader Star Wars universe. These can be inferred from the below word clouds for each:

Fig 5: Positive Tweets

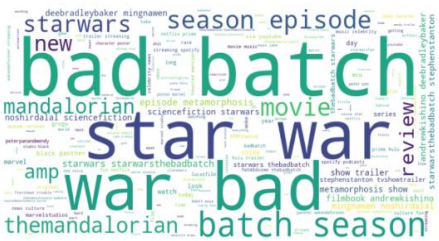

Fig 6: Negative Tweets

Fig 7: Neutral Tweets

B. Top Tweets and Most influential Tweet analysis:

The most engaging or influential tweets were determined by summing the number of retweets and favourites each tweet received. Subsequently, the top 10 most influential tweets were identified based on this criterion. Among them, 5 were attributed to the user **'985916593065476096'**, who also authored the top-ranked tweet. The top tweet was:

> Only 👏 👏 👏 for #jhope. j-hope IN THE BOX, a documentary special, is now streaming on #DisneyPlus! https://t.co/pNq

This tweet garnered **13,539 retweets** and **35,065 favourites**, marking it as the most impactful. Posted on February 26th, 2023, it highlighted the then recently released documentary film featuring BTS member Jhope, capitalizing on the global popularity of the KPOP group. Notably, the author boasts the highest follower count, enhancing the tweet's reach. Encouraging the viewership of the documentary, the tweet exemplifies effective influencer engagement. Other top tweets covered diverse topics, including discussions on 'Mandalorian' and 'Vision Quest', all reflecting a predominantly positive sentiment.

C. Topic Modelling for Disney Plus tweets:

The analysis of topics discussed during the specified timeframe revealed significant insights into prevalent trends. Utilizing KMeans with TF-IDF, NFM with TF-IDF, and LDA methods, three prominent themes emerged.

By using Kmeans these topics were found:

Topic 0	'marvel', 'movie', 'amp', 'plus', 'series', 'marvelstudios', 'streaming', 'trailer', 'animation', 'new'	The discussion revolving around Marvel Studios and new movies by Marvel.
Topic 1	'themandalorian', 'starwars', 'mandalorian', 'grogu', 'season', 'starwars https', 'themandalorian starwars', 'new', 'pedropascal', 'thisistheway'	The discussion on the Star Wars movie and 'The Mandalorian series, highlights both recent releases and the popularity of key characters and cast members like Grogu and Pedro Pascal.
Topic 2	'news', 'prime hulu', 'fun netflix', 'hulu trailers', 'culture fun', 'movies music', 'music celebrities', 'netflix prime', 'trailers streaming', 'spotify podcasts'	The discussion focused more on different streaming platforms like Netflix, prime, Hulu, Spotify, etc.

However, employing LDA revealed topics with more detailed insights, presenting some variations compared to KMeans & TF-IDF:

Topic 0	marvel, day, hotstar, disneyplushotstarth, peter, pan, peterpanandwendy, marvelstudios, animation, acc	On 1st March Marvel launched the trailer of the anticipated movie Peter Pan and Wendy and thus there was a decent volume of tweets directed to this topic.
Topic 1	trailer, movie, hulu, news, streaming, netflix, music, fun, culture, podcasts	The discussion focused on different streaming platforms like Netflix, prime, Hulu, Spotify etc.
Topic 2	themandalorian, starwars, mandalorian, season, grogu, episode, new, watch, mando, amp	The discussion about the star wars, the new season of The Mandalorian series, and its popular characters like Grogu.

While both KMeans and LDA identified major themes such as Marvel, Mandalorian, and streaming platforms, LDA's deeper insight, particularly regarding the inclusion of "Peter Pan and Wendy," underscores its value in uncovering more specific and nuanced topics within Twitter discussions. The word cloud for overall discussion made throughout the time range showed **"Mandalorian"** and **"Star War"** being discussed the most.

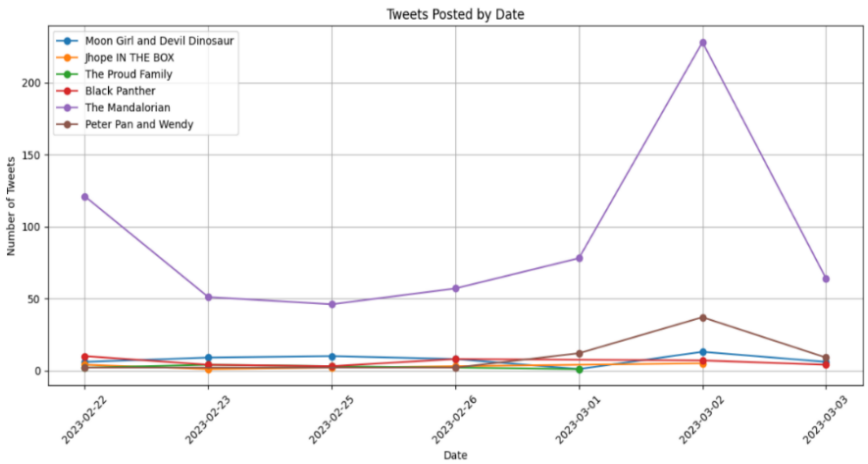

Fig 8

There were **6 major releases during the period of 15th February to 3rd March**. Fig 8, illustrates the no. of tweets made for all these major releases. It is clear that "The **Mandalorian" has stayed in hype** throughout. With the **'Peter Pan and Wendy'** trailer drop on 1st March, there is a **visible spike** in the no. of tweets for the topic on 2nd March. Also, the average sentiment about discussing these major releases has stayed positive.

As mentioned earlier, **March 2nd witnessed the highest volume of tweets** for that day, totalling 587, thus, a detailed analysis was conducted to explore the prevalent topics of discussion. These discussions primarily revolved around the same top three topics which were being discussed throughout the timeframe with **topics 1 & 2 related to streaming content and "The Mandalorian"** series dominating the conversation, indicating significant interest and engagement among users throughout the day. Furthermore, a subset of **47 tweets directly mentioned the brand** '@DisneyPlus', '@Disneyplus', or '@disneyplus', reflecting the platform's active engagement and involvement in the ongoing discussions.

D. Segmenting users using clustering:

To ensure effective clustering, a new dataframe was constructed by aggregating features such as tweet count, active duration, retweets, favourites, hashtags used, user sentiment score, and follower counts. Later, a high correlation (0.98) between the number of retweets and favourites was found thus the removal of the former from the analysis was done. Further PCA was applied, resulting in the creation of three principal components, with the first two components explaining 87.8% of the cumulative variance. Notably, the feature **"no_of_followers"** exhibited the **highest absolute weight of 0.9396**, followed by **"no_of_favourites"** with a weight of **0.7**. Subsequently, all users were segmented into three clusters using the KMeans algorithm.

Cluster 1:

This cluster has **356 users** representing a group of Twitter users with a **low number of followers** (mean: 25.24) posting a mean of 1.87 tweets, and active for about 1.28 days on average. They tend to receive a zero to a very low number of retweets and favourites, even though on an average they use 12 hashtags. Thus, the **engagement rate** here is **very low**.

Cluster 2:

This cluster has **548 users** representing a group of Twitter users characterized by **moderate engagement levels**. On **average**, they have around **1,046 followers** and post approximately 1.3 tweets per day. Their tweets receive minimal engagement, with an average of 1.2 retweets and 2.8 favourites per tweet. They use on average 7 hashtags. And have an engagement rate of 1.362% engagement.

Cluster 3:

Cluster 3 has **169 users** representing a group of users with a **high number of followers and average 6.8 tweets**, active over a period of about 2.27 days on average. They tend to receive an average of 124.55 retweets and 538 favourites, with a relatively high count of hashtags used (40 hashtags). The engagement rate for this cluster is around 8.65%, indicating a **high level of interaction** with their content as compared to the other 2 clusters.

5. Micro-influencers recommendation:

Further refinement was conducted on the previously generated dataframe to exclude authors with a follower count below 300,000, resulting in the identification of the top 9 influential authors. From this selection, the micro-influencer was selected based on a combination of follower count and overall engagement rate. One such author, boasting approximately **5.2 million followers** and an **engagement rate of 6.48%,** emerged as the micro-influencer, identified by the user ID **'985916593065476096'**. This individual has contributed a total of 25 tweets, amassing 15,453 retweets and 51,165 favourites across all posts, indicating substantial engagement with their content. Notably, this influencer's most impactful tweet pertained to be about the documentary film 'Jhope IN THE BOX'. Furthermore, upon examining the top tweets across all 7 active days, this author claimed the top spot on 4 out of 7 occasions. The author predominantly focused on discussing three major topics: **Topic 0 (Mandalorian), Topic 1 (Wakanda Forever) and Topic 2 (The crossover series),** refer Fig 9.

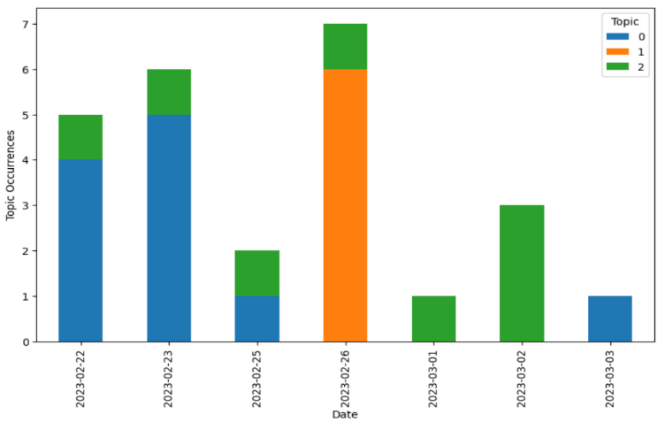

Fig 9

In the topics found, the mention of 'Chapter 17' likely refers to the title of the first episode of Season 3 of 'The Mandalorian', which debuted on March 1st. Overall, the sentiment of the influencer is very positive towards the brand. An analysis of the language utilized in the author's tweets indicates a consistent encouragement for viewers to engage with the new shows highlighted on the Disney Plus platform. For instance:

Keep your eye on the ball. 🏀 #TheCrossover, an Original series, is streaming April 5 on #DisneyPlus. https://t.co/bWGO7BR3vW

Upon further investigation, it was found that this influencer is the **official Twitter/X handle of Disney Plus**. This clarifies why the tweets consistently encouraged viewers to watch newly released content.
As the identified account belongs to **Disney Plus**, it disqualifies it as a micro-influencer. As per definition, Micro-influencers typically have follower counts ranging from 10,000 to 100,000. After sorting based on this criterion, and further prioritization based on the engagement rate, the **author** with ID **21113724** emerged as the **second potential influencer**. Several factors contributed to this assessment. The author is a verified user and was active on Twitter for 4 out of 7 days and posted five tweets. The sentiment analysis indicates an overall positive sentiment. Additionally, the author has 61,728 followers. Despite posting only five tweets, the author used 14 hashtags to expand the reach and garnered 1152 favourites and 135 retweets, indicating significant engagement with the content. Furthermore, the author's language appeared to encourage viewership of content on Disney Plus. For example:

Check out the awesome #TheMandalorian outdoor in Los Angeles. 3 days to go! #StarWars #DisneyPlus https://t.co/4R47v5dMYP

Comparison of micro-influencer with the most active Twitter user in the given time frame:
The user who made the maximum tweets in this time frame was author- '**1534870262679748609'** from Los Angeles, CA location with 126 tweets. The author has 24,793 followers and has been active on all 7 days. However, the engagement to the tweets received is very low, with only 55 favourites and 5 retweets for all 126 tweets made. If further information about when the person joined twitter was available then we can get a better idea that if the person is a new user and based on his online activity the person has a good prospect of being an influencer.

6. Conclusion:

The analysis of Twitter data focusing on Disney Plus provided valuable insights into brand sentiment, influencer engagement, and prevalent topics of discussion. Overall, the sentiment towards the brand was predominantly positive to neutral, with only a small fraction of tweets expressing negativity. Key topics of discussion revolved around popular shows like "The Mandalorian" and Marvel Studios, as well as discussions about various streaming platforms. The identification of a micro-influencer, along with the segmentation of Twitter users into three clusters, further elucidated the dynamics of user engagement and behaviour within the online community.

Insights:

- Sentiment: Overall, sentiment towards Disney Plus on Twitter was mostly positive, with occasional negativity.
- Influencers: A micro-influencers identified had a significant impact, driving positive discussions around Disney Plus content. If more information related to when they joined the platform was available, better insights can be achieved.
- User Segmentation and engagement: Twitter users clustered into distinct segments, offering insights into engagement levels and behaviour. It is mostly observed that the people tweeting most are using many hashtags however receiving lesser engagement on the contrary people using fewer hashtags and posting less have secured more retweets and favourites.
- Topics: Discussions centered on popular shows like "The Mandalorian" and Marvel Studios, emphasizing the competitive streaming landscape.

Recommendations:

- Influencer: Utilize micro-influencers to boost brand visibility and engagement as many users are tweeting actively and trying to increase reach using hashtags and brand tags.
- Content Strategy**:** Tailor content to match prevalent topics and user preferences.
- Community Engagement: Foster a sense of community to enhance brand loyalty.
- Monitoring: Continuously monitor discussions and sentiment for proactive response strategies.