**BRAND ANALYSIS – NETFLIX**

1. **EXECUTIVE SUMMARY**

The task at hand is to conduct an exploratory analysis of the public's perception of Netflix through Twitter data and identify a micro-influencer for the company to engage with. The report describes an analysis of a dataset containing information about tweets related to Netflix. The analysis includes data cleaning and preprocessing, time-series analysis, sentiment analysis, hashtag analysis, top locations, counting of top hashtags and mentions, topic modeling with NMF, and micro-influencer identification. The dataset contains 2649 rows and 15 columns, with information about 2649 tweets from 1891 unique users.

Sentiment was fairly positive with 1132 positive tweets, 558 negative tweets and 959 neutral tweets. Number of tweets peaked on Feb 23rd, Feb 26th and Mar 3rd due to Netflix Original movie ‘The Strays’ gaining popularity and Netflix taking action against multiple users sharing same account. Top 8 topics discussed in tweets include giveaway for premium account/contests to win accounts; documentaries including “Seaspiracy,” “Cowspiracy,” “Dominion”; TV show “Outer Banks” new season; technical issues with website/URL; production of animated film/TV show based on source material; auction/sale tickets/products; TV show episodes including “Ghost”; private 4K service likely adult content.

A micro-influencer with potential to collaborate with Netflix identified as Author ID: 36008477 who has significant follower count (84,671) indicating large reach for content; notable engagement rate; active Twitter presence; relevance to Netflix-related topics; positive sentiment in tweets; ease of collaboration.

1. **APPROACH BREAKDOWN**

* **Data cleaning and preprocessing:** The dataset is loaded into a pandas dataframe, and necessary libraries are imported to the code. To ensure that the data is cleaned and preprocessed, a function named preprocess\_text is defined. This function eliminates URLs, mentions, hashtags, special characters, and punctuation from the text column of the dataframe. The apply() method is used to apply the preprocess\_text function to the text column. A new column named tokens is created by tokenizing the preprocessed text using the word\_tokenize() function from the nltk library. Stop words are removed from the tokens column using the stopwords module from nltk. The filtered tokens are then stored in another new column named filtered\_tokens. The WordCloud function from the wordcloud library is used to generate a word cloud of the most frequently occurring words in the filtered\_tokens column.
* **Time-series analysis**: To perform time-series analysis on the dataset, the created\_at column of the dataframe is converted to a datetime format, and it is set as the index. Resampling of the data is carried out to daily frequency using the resample() method, and the number of tweets per day is counted. Using the plot() method from matplotlib, the number of tweets over time is plotted and the common topics discussed are displayed.
* **Sentiment analysis:** The SentimentIntensityAnalyzer class is imported from the nltk.sentiment.vader module, and an instance of the class is initialized. The text column of the dataframe is then processed with the analyzer instance to calculate the sentiment score for each tweet. The calculated scores are stored in a new column called sentiment. The hist() method from matplotlib is used to plot the distribution of sentiment scores.
* **Hashtag analysis:** A function named preprocess\_hashtags is defined to clean and preprocess hashtags in the hashtags column of the dataframe. However, the function is defined multiple times with the same code, which is unnecessary. The code does not appear to use this function for any further analysis.
* **Top locations:** The value\_counts() method is used to count the number of tweets from each location, and the top 10 locations with the most tweets are displayed. However, the code does not use this information for any further analysis.
* **Preprocessing of hashtags:** The preprocess\_hashtags function is applied to the hashtags column of a DataFrame called netflix to clean up the hashtags and remove any special characters or spaces.
* **Vectorization of hashtags:** The CountVectorizer function from the sklearn library is used to transform the cleaned hashtags into numerical features that can be used for machine learning algorithms.
* **Counting of top hashtags:** The frequency of each hashtag is counted, and a DataFrame called word\_counts\_netflix is created that shows the top 15 most frequent hashtags in the dataset.
* **Counting of mentions:** The number of times the word "netflix" appears in the text column of the DataFrame is counted, and the total number of mentions is printed.
* **Topic modeling with NMF:** The TfidfVectorizer function from the sklearn library is used to transform the text column into numerical features. Non-Negative Matrix Factorization (NMF) is then applied to find 10 topics within the text data.
* **Micro-influencer identification:** The netflix DataFrame is filtered to only include authors who are verified and have posted over 10,000 statuses. The remaining authors are sorted by sentiment, status count, and engagement rate, and the top 5 authors with the highest engagement rate are selected.
* **Selection of the final micro-influencer:** The influencer with the highest engagement rate is selected, while also considering their statuses count and sentiment, to finalize one influencer to work with.

1. **DATA DESCRIPTION**

|  |  |
| --- | --- |
| Column Name | Description |
| author\_id | Unique ID of the Twitter user who authored the tweet |
| statuses\_count | The total number of tweets posted by the Twitter user |
| location | The location of the Twitter user as indicated in their profile |
| verified | A Boolean flag indicating if the Twitter user has a verified account |
| followers\_count | The number of Twitter users who follow the author of the tweet |
| friends\_count | The number of Twitter users the author follows |
| listed\_count | The number of public Twitter lists the author is a member of |
| favourites\_count | The number of tweets the author has marked as a favourite |
| tweet\_id | Unique ID of the tweet |
| text | The text content of the tweet |
| created\_at | The date and time when the tweet was created |
| lang | The language in which the tweet was written |
| retweet\_count | The number of times the tweet has been retweeted |
| hashtags | A list of hashtags (if any) included in the tweet |
| favorite\_count | The number of times the tweet has been marked as a favourite |

Data range: 2649 rows × 15 columns

Date range: February 23 to March 3, 2023.

Number of tweets: 2649

Number of unique users analyzed: 1891

1. **ANALYSIS**

Netflix was highly prevalent and Netflix was mentioned 565 times in these tweets (using @).

Here is a **word cloud** that includes the most commonly tweeted words:

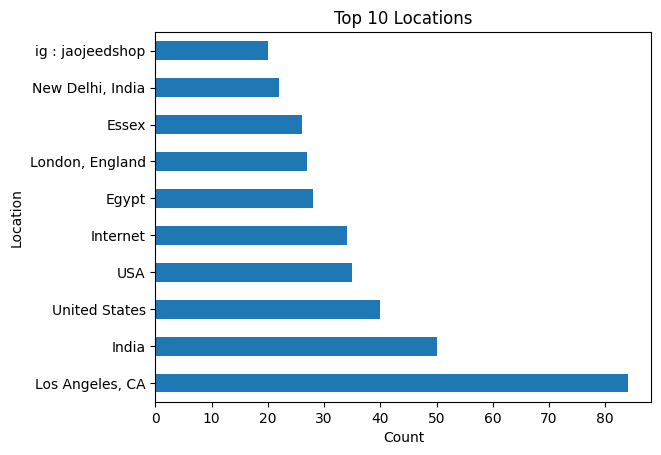


* There was an unusual trend where old documentaries such as ‘cowspiracy’ and ‘seaspiracy’ were being discussed.
* The word "amp" often appears because it is a HTML entity used to represent the ampersand character (&) in text. In many cases, when people share links or hashtags that contain "&" symbol, the Twitter platform automatically replaces it with "&", which can then appear frequently in the text of tweets. Therefore, when tokenizing tweets, the "&" entity is treated as a separate word, which can lead to a high frequency of the word "amp".

|  |  |
| --- | --- |
| Word | Count |
| netflix | 2670 |
| movies | 153 |
| outerbanks | 111 |
| streaming | 108 |
| hulu | 94 |
| disneyplus | 89 |
| movie | 89 |
| music | 87 |
| news | 84 |
| film | 81 |
| drivetosurvive | 78 |
| prime | 67 |
| tv | 64 |
| fun | 64 |
| spotify | 62 |

* ‘Outer banks’ was one of the most discussed shows in the tweets.
* Another evident discussion was related to the series ‘drive to survive’.

The location section was broken and rendered unusable for geospatial analysis. However, by plotting the top recorded locations of tweets, it was evident that Los Angeles, CA and India contributed majorly as seen below:



**Topic Modelling Analysis:**

Top 8 topics that were discussed in the tweets are,

*Topic #1*

['streaming', 'just', '20', 'win', 'participate', 'giveaway', 'premium', 'free', 'account', 'netflix']

This topic seems to be related to a giveaway for a premium account and contests to win accounts. This should be taken seriously and eliminated as it poses the risk of fraud in the name of Netflix subscriptions.

*Topic #2*

['en', 'te', 'recomiendo', 'documentales', 'seaspiracy', 'changers', 'cowspiracy', 'dominion', 'health', 'game']

This topic appears to be related to documentaries, including titles like "Seaspiracy," "Cowspiracy," and "Dominion."

*Topic #3*

['cast', 'episode', 'dont', 'drive', 'survive', 'new', 'streaming', 'banks', 'outer', 'season']

This topic seems to be related to a TV show called "Outer Banks" and its new season.

*Topic #4*

['possible', 'website', 'brand', 'url', 'vt', 'ip', 'detected', 'clone', 'address', 'follow']

This topic appears to be related to some technical issues with a website and its URL and requires attention as this can hamper the viewing experience.

*Topic #5*

['working', 'film', 'animated', 'based', 'adaptation', 'project', 'giant', 'guillermo', 'del', 'source']

This topic appears to be related to the production of an animated film or TV show based on some source material.

*Topic #6*

['ebay', 'belowgrab', 'auction', 'gp', 'sold', 'link', 'baby', '2023', 'day', 'ticket']

This topic seems to be related to some auction or sale of tickets or products.

*Topic #7*

['episodes', 'got', 'really', 'today', 'finished', 'yall', 'ghost', 'just', 'im', 'watching']

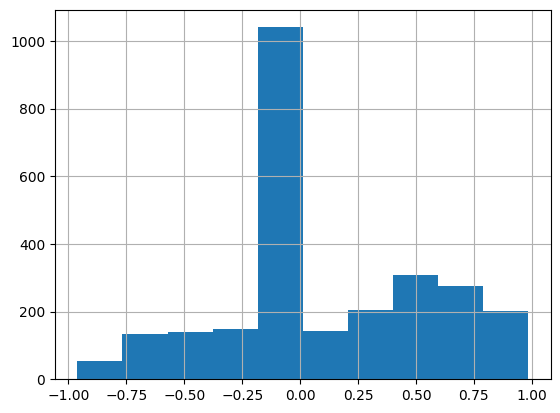
This topic appears to be related to TV show episodes, including "Ghost." It seems to be the ending of the season and sparked quite the discussion on twitter.

*Topic #8*

['private', 'th', 'dm', '4k', 'ตฟลกซราคาถก', 'ตฟลก', '30', 'days', 'amp', 'ตฟลกซ']

This topic appears to be related to a private 4K service. It is likely adult content and thus actions must be taken to ban these tweets in Netflix conversations.

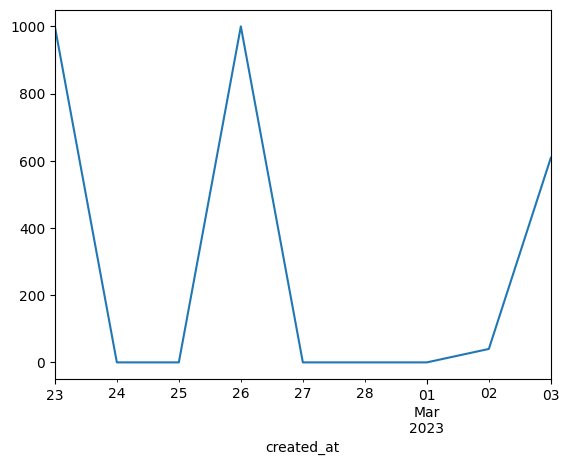
The **sentiment was fairly positive** as seen in the graph below:



Number of **positive** tweets: 1132, number of **negative** tweets: 558 and number of **neutral** tweets: 959

**Time Series Analysis:**

The number of tweets peaked on 3 days (Feb 23, Feb 26, Mar 3) as seen in the line plot below:



It is likely due to ‘The Strays’, Netflix Original movie gaining popularity. Another reason to spark a discussion was the fact that Netflix is going to take action against multiple users sharing the same account. On 26th Feb alone there were 1000 tweets and the tweets were mainly centered around the movie ‘We Have A Ghost’ and also an alarming number of tweets regarding giving out free Netflix premium accounts and some trusted reseller.

1. **MICRO-INFLUENCER RECOMMENDATION**

The micro-influencer who has the potential to collaborate with Netflix is Author ID: 36008477. This Influencer is verified by Twitter and has 84,671 followers with a status count of above 10,000 indicating active participation on twitter.

|  |  |  |  |
| --- | --- | --- | --- |
| Tweets | Hashtags | Engagement Rate | Sentiment |
| with both amp now hosting adaptations of s works its time one of them helped gaiman and others to of | netflix amazon neilgaiman rickywhittle finishthestory americangods opinion | 0.0012637148492399995 | 0.0 |
| icymi with both amp now hosting adaptations of s works its time one of them helped gaiman and others to of | netflix amazon neilgaiman rickywhittle finishthestory americangods opinion | 0.0005669001192852334 | 0.0 |
| and amp dark 1899 are adapting iv amp s from | netflix baranboodar jantjefriese jamestynion wertherdelledera somethingiskillingthechildren boomstudios | 0.0003425021554014952 | 0.0 |
| showrunner took to twitter to share more bts goodies with the fans in appreciation of the campaign | warriornun simonbarry savewarriornun netflix appletvplus halobearers | 2.3620838303551394e-05 | 0.6705 |

(amp is probably a media file and thus will be ignored here)

Here are some strong reasons to choose Author ID 36008477:

* **Follower count:** The influencer has a significant follower count of 84,671, indicating a large reach for their content.
* **Engagement rate:** While the influencer has only tweeted a few times recently, their engagement rate is still notable, with one tweet having an engagement rate of 0.0012637148492399995. This suggests that their followers are highly engaged with their content.
* **Active Twitter presence:** This person tweeted three times in just over a week, which shows that they are active on the platform and engaged with their audience.
* **Relevance:** The influencer has tweeted about Netflix and Amazon hosting adaptations of Neil Gaiman’s works and have used hashtags such as #netflix, #amazon, #neilgaiman, #rickywhittle, #finishthestory, #americangods and #opinion. Topics related to Netflix have been highly recurring in these tweets. This indicates that the influencer has an interest in these topics and potentially has followers who share that interest, which could make this influencer a good fit for collaboration.
* **Positive sentiment:** One of the tweets has a sentiment score of 0.6705, indicating a positive tone.
* **Ease of Collaboration:** Judging based on the nature and content of tweets and hashtags, the influencer seems to be an individual and not a group of individuals and taking into account the number of followers, it wouldn’t cost much for Netflix for a partnership with this influencer and might end up being a good future investment also as the influencer has audience who read content related to Netflix.

1. **CONCLUSION**

In conclusion, this brand analysis report provides valuable insights into the public perception of Netflix through Twitter data analysis. Sentiment analysis shows that the majority of tweets about Netflix are positive, indicating that Netflix has a strong brand image and customer satisfaction. The most common topics discussed in tweets about Netflix include documentaries, TV shows, technical issues, production of animated film/TV shows, TV show episodes. The top hashtag used in tweets about Netflix is #Netflix, followed by #Seaspiracy, #Cowspiracy, #Dominion, #OuterBanks, and #Ghost. The identified micro-influencer, Author ID: 36008477, has a significant follower count (84,671), notable engagement rate, positive sentiment in tweets, and is relevant to Netflix-related topics. They would be an excellent choice for collaboration to promote Netflix content to a wider audience.

Documentaries are a popular topic among Twitter users discussing Netflix. By investing in documentary content, Netflix can continue to attract viewers and generate buzz on social media. Netflix should also monitor and remove any inappropriate content on their Twitter handle to maintain their brand image and ensure customer satisfaction.