Brand Analysis using Twitter Data

By: S M Sazid Islam

Executive Summary

This report presents an exploratory "snapshot" analysis of the public's view of Rockstar games company as seen through the lens of social media (twitter). To conduct this report, tweets are extracted from twitter with keyword RockstarGames in the span of a week. Rockstar games is an American video game publisher based in New York city. The main objective of the report is to analyze the brand's relationship with people on social media. The analysis was performed using the Jupyter Notebook (anaconda3) python programming language. Sentiment analysis, topic modelling, application of classification modelling and finally micro-influencer is recommended based on the tweet interactions.

Approach

The methodology of the study has been explained in this section. Two python Jupyter notebooks have been used to perform the study successfully. In the first python file (Notebook-1), tweepy is installed first using pip command: pip install tweepy and imported for extracting the tweets in the notebook as well as Twitter API (Application Programming Interface) is used. In order to extract the data, the API keys were used which has been copied from the twitter development application. The keys allow us to get the tweets from twitter. Next, the date has been set to make it time bound and finally after the extraction the tweets are saved into a csv file with the name Rstar_1.csv.

The extracted tweet data files are now uploaded in a new Jupyter notebook (Notebook-2) to perform the analysis. The files consist of many null values which are cleaned and prepared for the analysis. Tokenization, Stemming and Vectorization are also performed to pre-process the data file and make it proper for the analysis. After the data has been properly cleaned, word cloud is formed to summarize the most frequent tweets with sentiments: positive, negative and neutral. Pie chart and bar chart are plotted to show the distribution of sentiments visually. Next, classification modelling is applied to understand the accuracy of the words identified. Lastly, the positive tweets are considered and the potential micro-influencer is identified.

Collection of Data

For data collection, the data is extracted using the python Notebook which had a number of null values. After dropping the null values, 2230 tweets are finally collected with 17 features were considered for the analysis. Tweets made in between 3rd to 5th of May 2022 are extracted and merged together. The tweets were collected from all around the world. The features of the tweets are listed below:

id- Twitter Author Id	name: Account user name	screen_name: unique name of twitter	verified: Verified account or not
description: Twitter account bio	location: location of the user	followers_count: number of followers of the account	friends_count: number of friends of the account
listed_count: public lists number of user membership	favourites_count: tweets liked by the user	statuses_count: number of tweets made by the user with retweets	created_at: the time at which the tweet was created
Text: the tweet content	retweet_count: times the tweet retweeted	hashtags: any hashtag used in the tweet	source: device by which the tweet was made
		favorite_count: times the tweet is liked	

Analysis and Description

This section includes the detail description of the analysis performed which includes pre-processing of the data, sentiment analysis, topic modelling and use of logistic regression to test the accuracy of sentiment analysis.

a) Pre-processing of data:

The pre-processing of the data is very important because the efficiency of the analysis depends on it. The tweets collected in a csv file is uploaded to a new python notebook (Analysis- Rockstar Games) for analysis. At first the tweets are examined and discovered that the tweets are not prepared to perform the analysis. Hence, a number of cleaning steps are performed like dropping null values, adding column names, removing website links (URLs) to other websites, tokenization which is the splitting of text into words, stemming- the process of reducing words with similar meanings and punctuations are also removed. Tokenization and stemming are performed using Natural Language Toolkit (NLTK) library. In addition, the text is converted into lower case and stop words like a, and, the etc. were removed. The retweets were also removed because there will be a lot of repeated tweets which will reduce the accuracy of the analysis. After cleaning mean is calculated and it is seen the mean of text after cleaning has reduced which indicates the successful cleaning of the text.

b) Sentiment Analysis:

The main goal of performing sentiment analysis is to classify the tweets accurately into different sentiment groups. To apply the sentiment analysis properly, the subjectivity and polarity of the tweets are also analyzed using TextBlob function. Polarity refers to the strength of an opinion which is measured on a range of +1 to -1. The tweets closer to +1 are positive tweets and the tweets closer to -1 are the negative tweets while 0 indicates the neutral tweets. The tweets are then grouped into 3 subgroups with the names ptweets: for positive tweets, negtweets: for negative tweets and neutweets: for neutral tweets. Next, the word clouds for most frequent positive, negative and neutral tweets are created to compile the most frequent top 100 sentiment words. As well as, the percentage of the tweets are also calculated which are visualized using pie and bar charts.

The word clouds are demonstrated below:

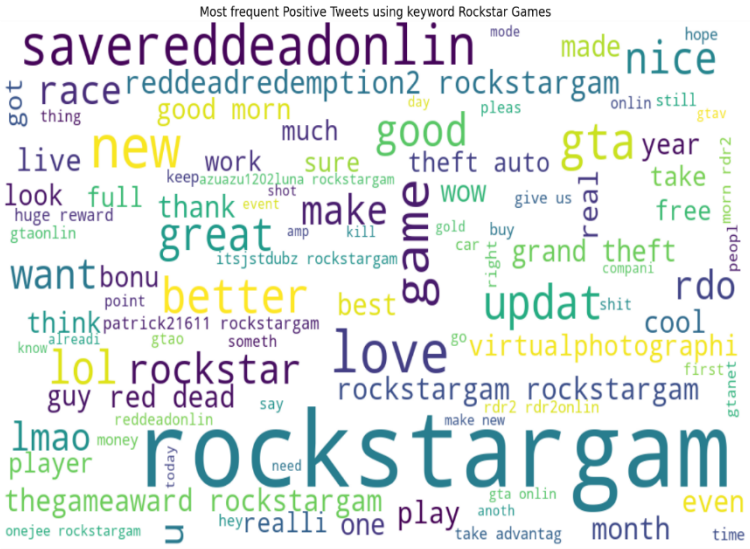


Figure 1: Positive words associated with positive tweets

Figure.1 shows the words people used to express positive emotions about Rockstar games. People expressed that they like games like “gta” (Grand Theft Auto) and “rdr2” (Red Dead Redemption 2) which are created by Rockstar Games. As well as, there is a petition going on in the social media to save online version of the game “Red Dead Redemption” for which people are sharing positive emotions about that “savereddeadonlin”. Moreover, there are positive words like “nice”, “great” and “love” which shows that the gamers are linking RockstarGames.

Figure 2: Negative words associated with negative tweets

Figure.2 shows the words twitter users used to express negative emotions about the brand. Here, it can be stated that people do not like the virtual photographing option, they have also expressed negative emotions about the topics that some other people have positive emotions like “red dead redemption”, “gta”. These are the options of the games that people do not likes or expects an upgrade.

Figure 3: Neutral words associated with neutral tweets

Figure.3 shows the neutral sentiments about the brand, people have shared neutral emotions about “gta5”, “rockstareditor”. Some have also shared neutral emotions about the petition “savereddeadonlin”. As well as, “thegameaward” -this is an award show where the best games get awarded and RockstarGames is also there, people have neutral emotions regarding this.

The distribution of positive, negative and neutral tweets is shown using pie and bar charts:

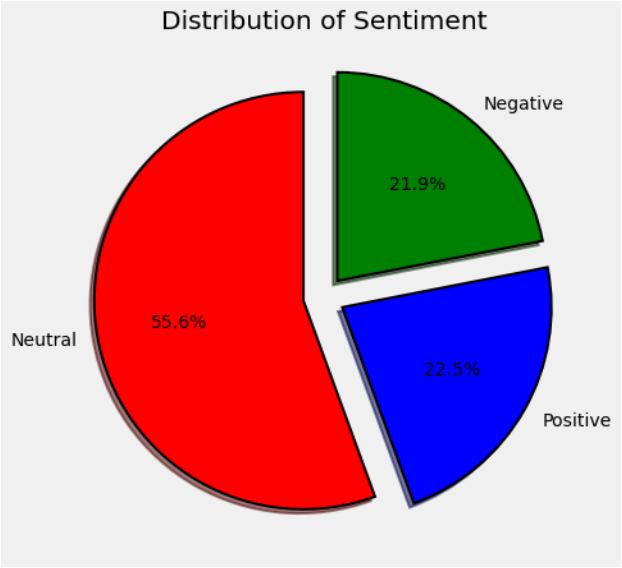


Figure 5: Distribution of sentiments in Pie-chart

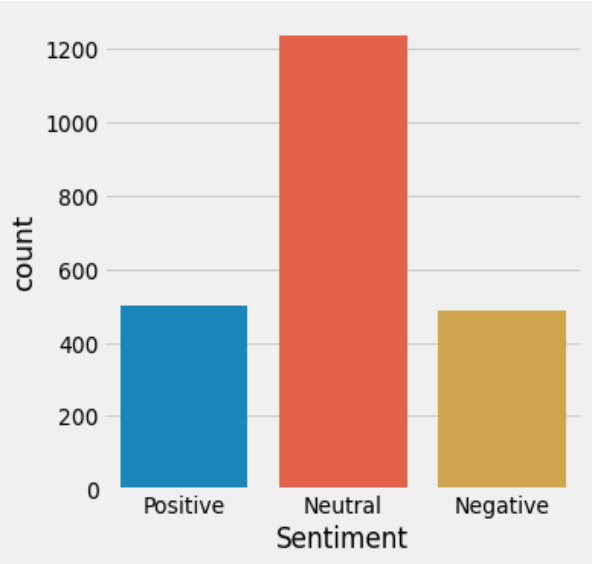
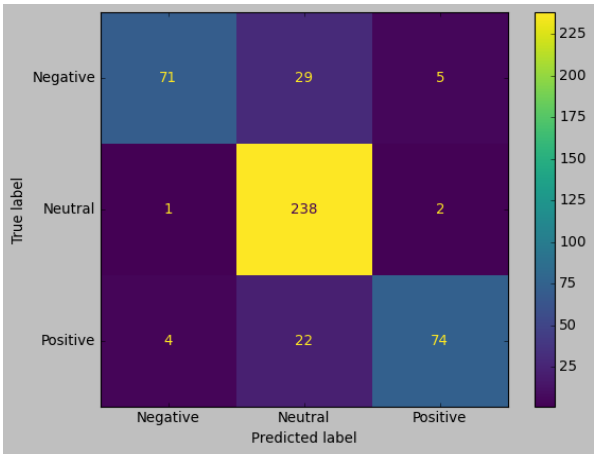


Figure 4: Tweet count of sentiments in bar-chart

Through analysis, we identified, 502 positive tweets, 489 negative tweets and 1239 neutral tweets.

The next step performed in the analytical process is the use of model to test the accuracy of the sentiment analysis. Logistic Regression is applied to the sentiments for which the clean text is converted to feature vector which is the conversion of words into vector with zeros and ones. The data is then divided into training and test set, 20% for testing and 80% for training. Next the accuracy of the test set is counted which is found as 85.87%. Next a confusion matrix is also plotted and the precision, recall and f1-score values are evaluated. Precision-TP/(TP+FP), this means that the proportion of points that model classify as positive are actually positives. Recall-TP/(TP+FN), this means the proportion of actual positives that are correctly classified by the model and f1-score is the harmonic mean of precision and recall. The confusion matrix is shown below with the evaluation metric values:



	Precision	Recall	F1-score
Negative	0.93	0.68	0.78
Neutral	0.82	0.99	0.90
Positive	0.91	0.74	0.82

Figure 6: Confusion-matrix display

Topic Modelling:

This is used to identify group of topics used by the people about Rockstar games. In order to identify the most used words and their frequency, the positive and negative words are separated into 2 different columns and visualized in tables. In order to perform topic modelling pre-processing steps like removal of URLs, converting letters into lowercase and also tokenization is important. The tables are shown below:

Positive Words	Count
new	59
gta	43
love	39
savereddeadonlin	36
good	35
game	34
play	32
race	29
updat	26

Negative Words	Count
game	213
dead	114
red	91
onlin	57
gta	54
rockstar	53
updat	35
savereddeadonlin	34
content	23

Based on topic modelling, top word like “new” was used the highest times to share positive emotions and “game” was used to share negative emotions. Next topic for expressing positive emotions is “gta”, this is name of a popular game which people likes a lot for which it is rated as the second most used positive word. Other positive tweets consist of words like “love”, “good” etc. This shows the relationship of RockstarGames with the people. There are also some highly highlighting negative topics which has been identified, one is “game”, this is a generic term that people has used to express negative emotions, as well as there is another word “savereddeadonlin”, this is used as both positive and negative word, this indicates that some people are in for the petition and some are against the petition. There is also a topic called “content” which shows that some people are not happy with the content of RockstarGames.

Micro-influencer recommendation

A micro-influencer can be an effective promotional tool for a brand in social media channels. A micro-influencer is any influencer who has followers from 5000 to 100,000 in a social media platform, for example. Twitter.

The micro-influencer was recommended based on the positive tweets only, the positive tweets are analyzed and the parameters like followers_count, friends_count are considered while identifying the micro-influencer for Rockstar Games as well as sentiment viz is also used to analyze the sentiments associated with the account. At first the followers_count of the top 10 twitter accounts are evaluated and from there accounts with screen name: “TwitchSIE”- 33098 followers, “sme_rt”- 29606 followers and “Retweelgend”- 12862 followers are selected as micro-influencers. Next, the friends_count of the top 10 accounts are evaluated and based on these two parameters, “sme_rt” is finalized as the recommended micro-influencer for future engagements because it has the highest number of friends_count of 8014 and also 29606 followers which is potential to be a micro-influencer. “sme-rt” is a popular gamers community with the origin in UK, the number of friends_count is also considered because this shows that the account

is trusted by many other users. For further analysis, using sentiment viz the sentiments associated with the account are visualized, the output of sentiment viz is shown below:

Figure 7: Sentiments related to "sme-rt"

Here it can be seen that people expressed positive emotions like “calm”, “relaxed”, “excited”, “contented” regarding the account “sme-rt”. Hence, this also adds up to the reason to recommend “sme-rt” as the recommended micro-influencer

Conclusion:

In a nutshell, the whole analysis is just a snapshot analysis of the public’s views about RockstarGames on the basis of the engagements made in Twitter. Sentiment analysis allowed us to understand the positive and negative feelings related to RockstarGames where 55.6% are neutral tweets, 22.5% are positive and 21.9% are negative. In addition, using Logistic Regression model, the accuracy of the sentiments is found to be 85.87% which indicates that prediction is good. As well as, using topic modelling the topics associated and frequency of the topics are also evaluated to identify the most used positive and negative words.

The twitter account – “sme-rt” is recommended as the micro-influencer because of its 29K followers and 8K friends added with the account. This is a gaming community which has a strong follower base and if RockstarGames engage with them they can promote their games better and build a stronger relationship with the people.

Lastly, this analysis can be further extended by considering geospatial locations to identify the emotions based on different areas and also by extracting dataset in the span of 1 month a more detailed and accurate snapshot can be achieved.