The Game awards 2022 Tweet analysis using Text mining

Benimam Abdelmadjid, G01

December 13, 2022

Contents

Introduction

Analysing textual data have become one the most interesting field this recent years. The exponential growth of online collected data from social media and web pages as well as newly discovered techiniques have played a major role in this development.

In this mini project i will analyse textual data about the game awards 2022 event using text-mining methods methods.
First section will contain a brief presentaion on the choosen topic(domain). Second section will show how the filtring process of textual data(tweets) was done. Third section will contain interpreations where important questions about positivity and popularity will be answered, additionally plots will be shown to visualize results properly.

Final section will be dedicated for conclusion, perspectives and possible future changes to improve the analysis.

1 The Game Awards

The Game Awards is an annual awards ceremony honoring achievements in the video game industry. Established in 2014, the shows are produced and hosted by game journalist Geoff Keighley, who worked on its predecessor, the Spike Video Game Awards, for over ten years. In addition to the awards, The Game Awards features premieres of upcoming games and new information on previously-announced titles.

1.1 Selection Process

The Game Awards has an advisory committee which includes representatives from hardware manufacturers Microsoft, Sony, Nintendo, and AMD, along with several game publishers. This committee selects around thirty influential video game news organizations that are able to nominate and subsequently vote on the video games in several categories. The advisory committee otherwise does not participate in the nomination or voting process. During the nomination round, each of the news outlets provides a list of games in several categories; games for the esports-related categories are chosen by a specific subset of these outlets. The committee compiles the nominations and selects the most-nominated games for voting by these same outlets. Prior to 2017, there were 28 industry experts and representatives that selected the winners, while the awards from 2017 onwards have used over 50 such experts. In 2019, non-English media publications were added to the jury. Winners are determined by a blended vote between the voting jury (90%) and public fan voting (10%) via social platforms and the show's website.

Generally, only games released before a specific date in November are eligible for being nominated in the year's awards. As the jury must make their nominations in the weeks prior to this date, this may leave some anticipated games that are scheduled for release just before that date to be underrepresented in the nominations, since the jury must go by pre-release review copies and not the final version. Any games releasing after the November deadline (which varies every year) are eligible for the next year's ceremony.

2 Filtring Tweets

This process can be partitioned into 2 part

2.1 Data aquisition

Twitter API pyton client **tweepy** was used to retreive 30000 tweets containing the hashtags "#TheGameAwards

2.2 data manipulation

Tweets with their propreties stored in json file were converted to a pandas dataframe.

DataFrame columns contains :

- `tweet_id`

- `created_at (publish date)`

- `text content of the tweet`

- `like_count`

- `retweet_count`

- `reply_count`

- `like_count`

- `quote_count`

- `author_id`

Figure 1: DataFrame illustration

3 Interpretation

3.1 Q1 : 5 principal regions

Although tweets region is not always activated by users around 600 regions were extracted from over 30000 tweets.

Figure2 Figure3 shows that majority of tweets coming from united states with over 300 tweets. india is the second with 50 tweets and canada is the last.
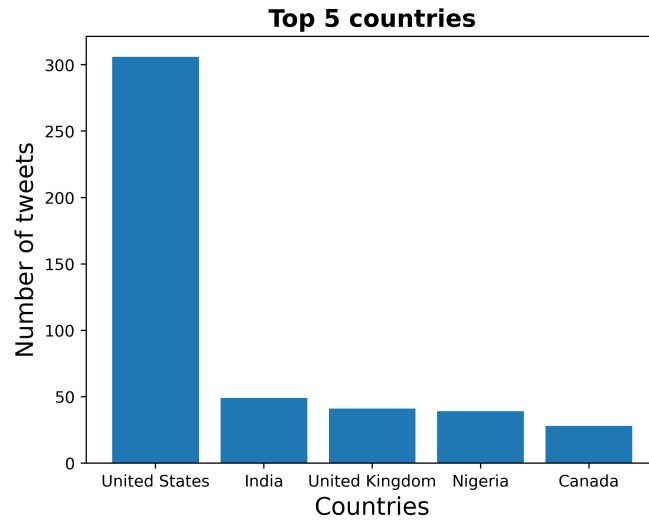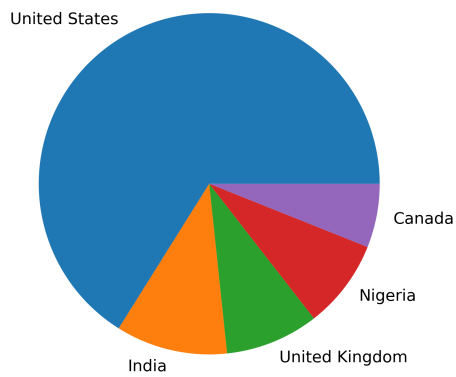

Figure 2: A bar chart representing top 5 tweets region

Figure 3: A pie chart representing top 5 tweets region

3.2   Q2 : Most popular Game

Twitter public metrics are usually 4:

- like count

- quote count

- reply count

- retweet count

From Figure4, Figure5, we can conculde that the most popular game/actor among participant is
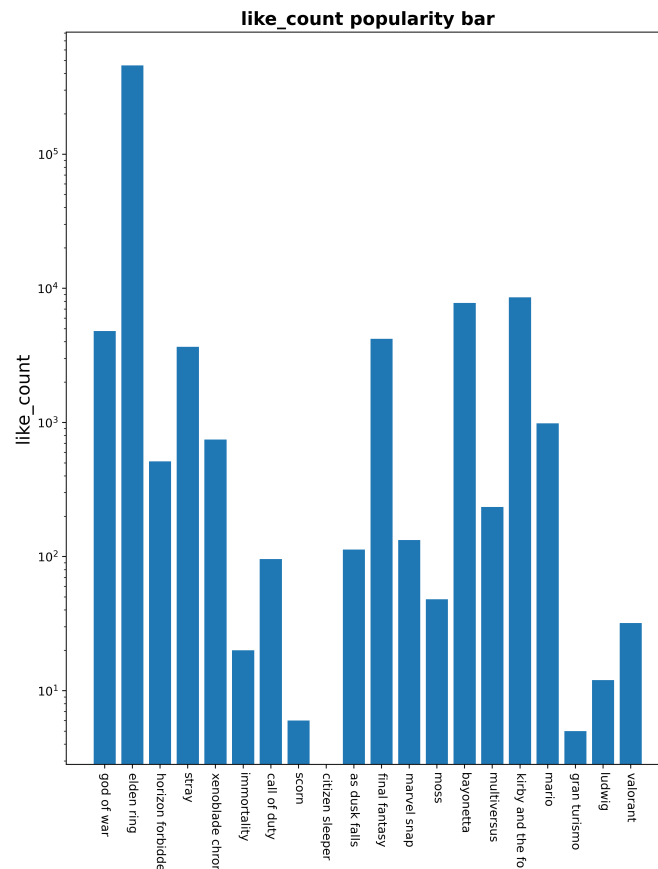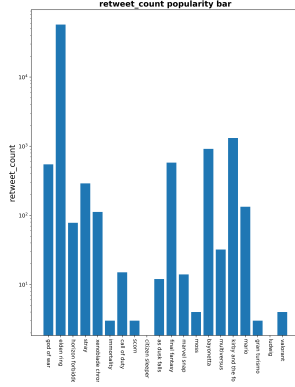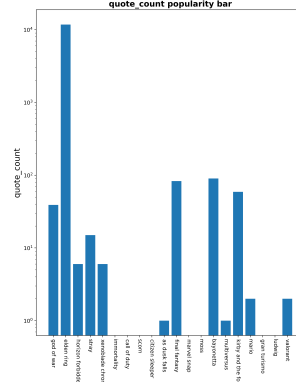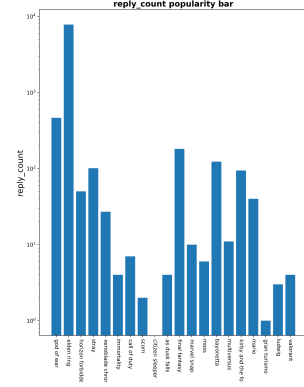**Elding Ring**

Figure 4: a bar plot representing each games/actor like count

(a) a bar plot representing each games/actor retweet count

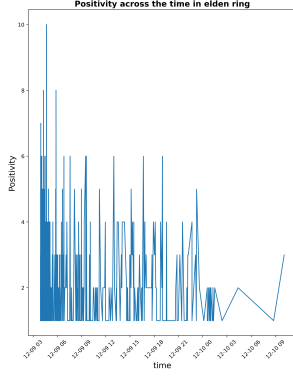(b) a bar plot representing each games/actor quote count

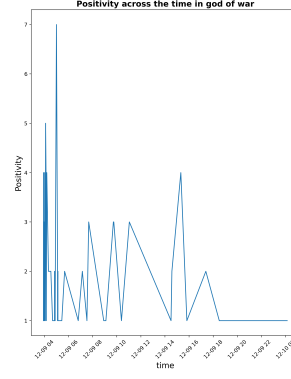(c) a bar plot representing each games/actor reply count

Figure 5: Counts Plots
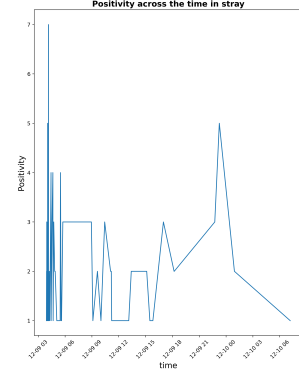
3.3   Q3 : Positivity Curve

Positivity rate was at it peak in the period between 09/12 03:00AM and 09/21 as Figure6 shows. where only 6 figures was choosen according to their size.
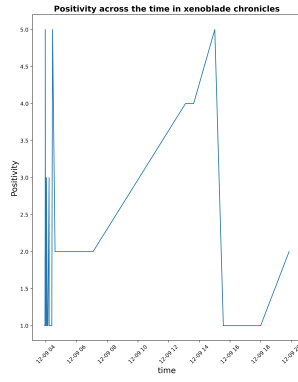
(a) positivity plot over time for elden ring
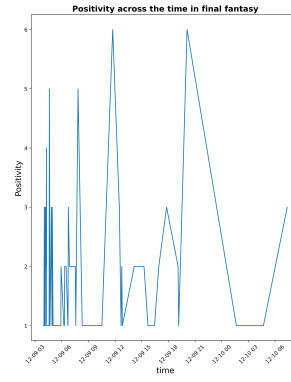

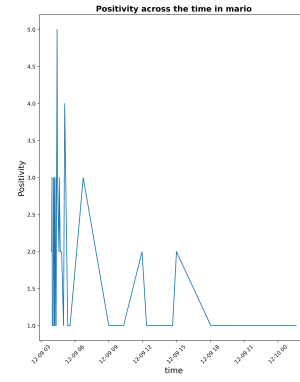(b) positivity plot over time for god of war

(c) positivity plot over time for stray

(d) positivity plot over time for xenoblade chronicles

(e) positivity plot over time for final fantasy

(f) positivity plot over time for mario

Figure 6: Positivity plots across the time

4 Conculsion

In this project a simple regular expression approach is used to determine tweets's polarities as well as related keywords.

Textual data have gone throgh a filtering processs in which only relevent data is preserved. Therafter stopwords, special characters, links, tags, and hashtags were removed from tweets. Forwardly relevent tweets to each game(subtopic) are stored seperately in csv format, and then plots of region, popularity, positivity were drawn using previous files.

Finally, Limits of this method have increased the number of false positive(FP) classification due to lack context in regular expression, However to improve text analysis a modern approach based on trainable models shall be used to decrease false positive data