

The Game awards 2022 Tweet analysis using Text mining

Benimam Abdelmadjid, G01

December 13, 2022

Contents

1	The Game Awards	4
1.1	Selection Process	4
2	Filtrring Tweets	5
2.1	Data aquisition	5
2.2	data manipulation	5
3	Interpretation	6
3.1	Q1 : 5 principal regions	6
3.2	Q2 : Most popular Game	7
3.3	Q3 : Positivity Curve	8
4	Conculsion	10

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 techniques 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 filtering process of textual data(tweets) was done. Third section will contain interpretations 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 retrieve 30000 tweets containing the hashtags ”#TheGameAwards

2.2 data manipulation

Tweets with their proprieties 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

	T	Text	T	Lang	Created_at	T	Author_id	T	Id	Retwee	Reply_coun	Like_count	Quote_coun
	1601322443662	ZIMBABWAY says stop sending that love love this period	en	en	2022-12-09T21:06:36.000Z	1599042545790734336	1601322443662	0	0	0	0	0	0
	1601187776891	Ziff Davis announcing layoffs at IGN hours before one of the biggest events of th	en	en	2022-12-09T12:11:29.000Z	239531641	1601187776891	0	0	0	0	0	0
	1601312371213	Yup I m pretty sure I m hyped	en	en	2022-12-09T20:26:34.000Z	1050179485281263616	1601312371213	0	0	0	0	0	0
	1601280517361	Yuji Naka seeing some random frilly hair kid in video game jail	en	en	2022-12-09T18:20:00.000Z	723018526663962624	1601280517361	0	0	0	0	0	0
	1601273013755	ysssss Croatia it s not over until you say it s over Get in	en	en	2022-12-09T17:50:11.000Z	3398641829	1601273013755	0	0	0	0	0	0
	1601337150490	ysssss Argentina Lucky win and lost a two goal lead	en	en	2022-12-09T22:05:02.000Z	3398641829	1601337150490	0	0	0	0	0	0
	1601332258547	Youre doing yourself a disservice if U have a amp havent downloaded amp pla	en	en	2022-12-09T21:45:36.000Z	1432520873022459904	1601332258547	4	0	0	6	0	0
	1601446601591	Your trusted hands for FG Funnels Leadkept Attract ROI Clixio projects	en	en	2022-12-10T05:19:57.000Z	1550209987825307648	1601446601591	0	0	0	0	0	0
	1601364715602	Your professional hacker is working on CasesSend in your Hacking Issues and gi	en	en	2022-12-09T23:54:34.000Z	1442643576568094720	1601364715602	2	0	2	2	0	0
	1601343317538	your odds haven t failed me and won t fail me that s why I will always bank on it	en	en	2022-12-09T22:29:33.000Z	1599772352186912768	1601343317538	2	1	2	2	0	0
	1601295616931	Your next favourite online RPG is coming soon and announced their collaborati	en	en	2022-12-09T19:20:00.000Z	1433846094165905408	1601295616931	0	0	3	0	0	0
	1601264810108	Your money is worthless sitting in a savings account Start your investment jour	en	en	2022-12-09T17:17:35.000Z	1592999236194471936	1601264810108	0	0	0	0	0	0
	1601299193980	Your Eagle is on The Live Experience you cant miss th DEC Tickets Available No	en	en	2022-12-09T19:34:13.000Z	1162995337281363968	1601299193980	1	0	2	0	0	0
	1601332122408	Your boys made it in best fits at	en	en	2022-12-09T21:45:03.000Z	394769648	1601332122408	4	11	69	0	0	0
	1601237461040	You will stand You will fail You will rise again Such are the challenges in life Nc	en	en	2022-12-09T15:28:54.000Z	1500770869043470336	1601237461040	1	0	1	0	0	0
	1601245183269	You ve earned some sleep after all those weeks planning Geoff Go rest now	en	en	2022-12-09T15:59:36.000Z	2183895764	1601245183269	0	0	0	0	0	0
	1601308125348	You ve always been there for me I found myself winning multiple times Am focu	en	en	2022-12-09T20:09:42.000Z	1335515526282489856	1601308125348	2	1	2	0	0	0
	1601249628061	You ve always been there for me I found myself winning multiple times Am focu	en	en	2022-12-09T16:17:15.000Z	1599451430838046720	1601249628061	2	1	3	0	0	0
	1601248644099	You ve always been there for me I found myself winning multiple times Am focu	en	en	2022-12-09T16:13:21.000Z	1463650662168907776	1601248644099	2	1	2	0	0	0
	1601248341509	You ve always been there for me I found myself winning multiple times Am focu	en	en	2022-12-09T16:12:09.000Z	1463650662168907776	1601248341509	2	1	4	0	0	0
	1601491936615	You too can also make money on affiliate marketing just hit me up on Whatsapp	en	en	2022-12-10T08:20:06.000Z	1537003743765483520	1601491936615	0	0	0	0	0	0
	1601316263166	You not cleaning unless you scrubbing them walls with bleach	en	en	2022-12-09T20:42:02.000Z	1596437671525556224	1601316263166	0	0	0	0	0	0
	1601260787078	You love to see it and IV lit up last night with a surprise performance of LILITH	en	en	2022-12-09T17:01:36.000Z	2806914710	1601260787078	192	31	2282	28	0	0
	1601239853156	You look lonelyI can fix that	en	en	2022-12-09T15:38:25.000Z	1534240790720958464	1601239853156	4	13	11	1	0	0
	1601237415943	You know why it s only the gif right	en	en	2022-12-09T15:28:44.000Z	1590018747263336448	1601237415943	0	0	0	0	0	0

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.

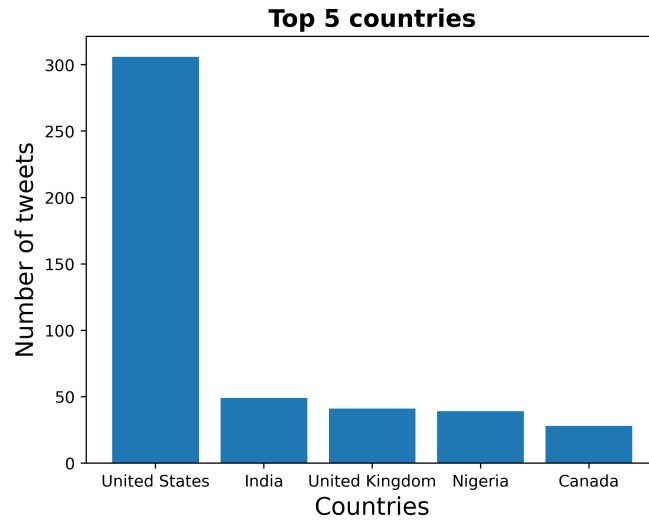


Figure 2: A bar chart representing top 5 tweets region

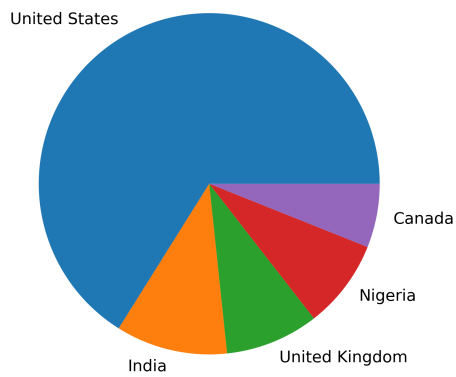


Figure 3: A pie chart representing top 5 tweets region

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 conclude that the most popular game/actor among participant is **Elding Ring**

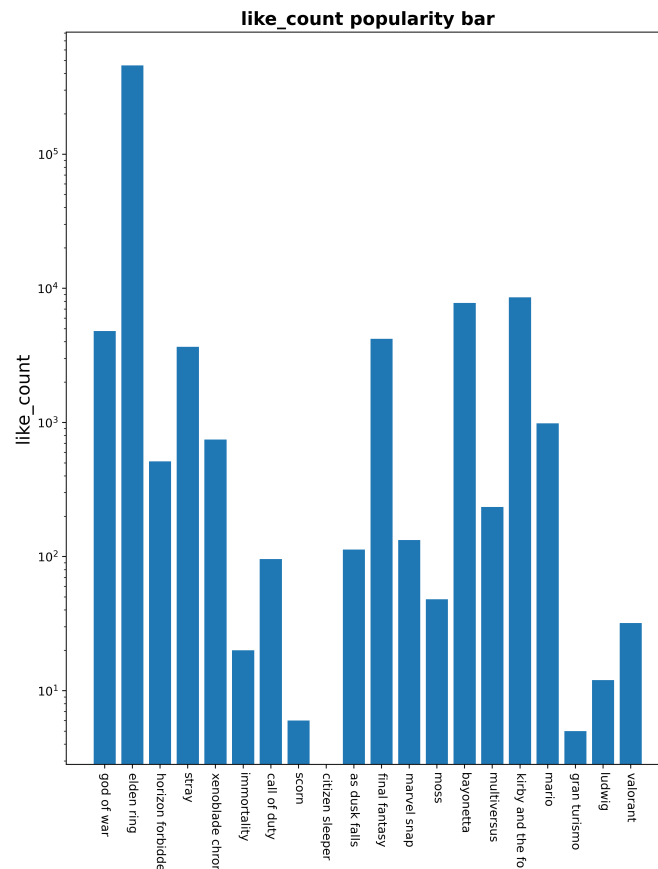
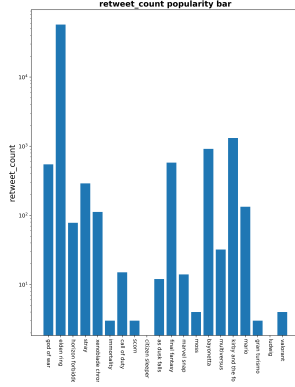
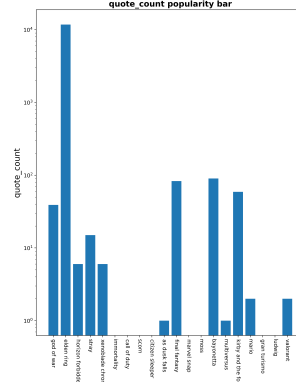


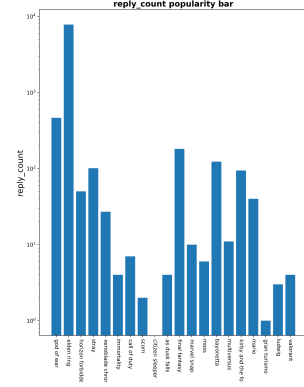
Figure 4: a bar plot representing each games/actor like count



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



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

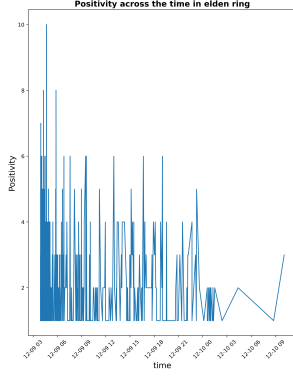


(c) a bar plot representing each games/actor reply count

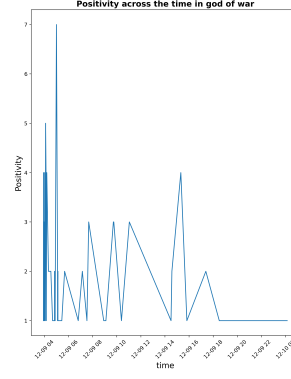
Figure 5: Counts Plots

3.3 Q3 : Positivity Curve

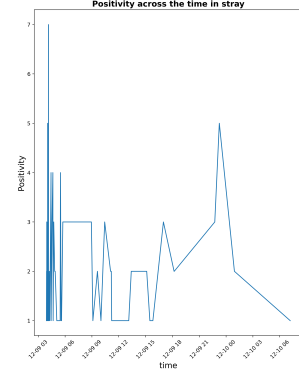
Positivity rate was at it peak in the period between 09/12 03:00AM and 09/21 as Figure6 shows. where only 6 figures was choosen according to their size.



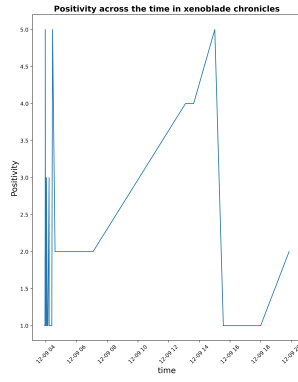
(a) positivity plot over time for elden ring



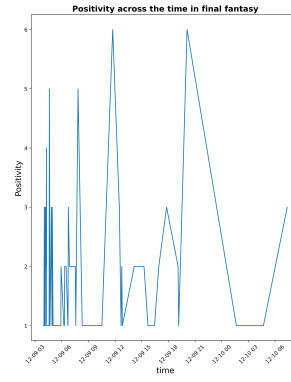
(b) positivity plot over time for god of war



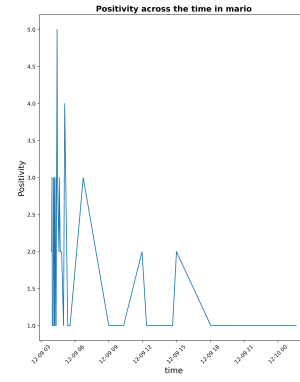
(c) positivity plot over time for stray



(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 through a filtering processs in which only relevent data is preserved. Ther- after 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