



Web Analytics

Group 6 PUBG Analysis

BYGB 7978 002

December 11, 2018

Hanwen Feng

Ruiqing Lyu

Yuwen Wu

Jeke Chen

Executive Summary:

Player Unknown's Battleground is a battle royale type game. In this game, a hundred players get ported into an island and fight to the death until there is only one winner. When it was released on March 27th, 2017, PUBG became one of the most popular games in the battle royale category. Starting from its releasing date till the beginning of January 2018, PUBG has gained over a high of 3,200,000 players in average monthly players. However, after that period, PUBG has experienced a severe decline in its user base. Also starting from January 2018, the number of players became continuously decreased to around only 418,665 players for now (SteamCharts).

There is no question that the decreased number of users would largely minimize Bluehole's probability of making a profit. There are many reasons that could lead to the decline in the number of players, but many of these are only speculative. Interestingly, we observed that January is the month when the first official patch of PUBG was released. We hypothesized that some of the gameplay updates released in the patches did not perform well with the game's user base. At the same time, some of the comparable games (such as Fortnite in our case), had increased in their number of users. In this report, we wish to use sentiment analysis as well as usage analysis to pinpoint exactly where PUBG had its downfall. We cross-referenced the usage data from Steamcharts with the patch dates of PUBG and picked out the top five months with the highest gain and loss percentage. We performed sentiment analysis of user comments in reaction to patches which were released within these months' range. By doing so, we can detect users' sentiments on the content of updates, and identify which action the users like and dislike.

On one hand, we derived term frequency pie charts based on our wordcloud of each patch review. From charts, we found there are many irrelevant words to the update content itself, such as "cheaters," "China" and "Fix."¹ The comments area under Steam review and on Reddit are open to everyone, so we guessed that players mostly comment about what they give special attention to, in other words, the points they are concerned with the most. Given what we have observed, we believe that infrequent bug fixes, various forms of cheating, as well as many Chinese players hacking, could be the reasons leading to the downfall of PUBG. On the other hand, we cross-referenced with Fortnite, which also has a battle royale play mode, to analyze its users' sentiment analysis. We found that there are more positive words within the wordcloud of Fortnite than that of PUBG. The differences are mainly indicated to playing modes or new maps. Consequently, we suggested Bluehole developers should definitely develop an effective method firstly to block cheaters or even detect gamers who use the cheater and ban them in the game lobby. Secondly, due to many Chinese players using cheats in the game of worldwide server, developers should take serious consideration with whether to build a separate server for Chinese region players only. Referring to the highlights of Fortnite, PUBG could create more play modes and also fix its bugs more often to satisfy its users.

¹ Please refer to Reference page to see further explanation for keywords

Introduction

The Video game industry is an industry that is highly competitive and highly fragmented. There are many different production studios building games to all types of people. Despite the high level of competition, niche game studios have found success because of how diverse the market is. Bluehole developers are a classic example of how a small gaming developer company gained massive popularity from their game Player Unknown's Battleground.

There are many different strategies for creating a successful game. The traditional style of gaming companies would be to release a new game every few years to build up their franchise. However, in recent years, studios have taken the approach of making games and continuing to develop them. By continuously patching and developing a game, companies effectively increase the longevity of the game, which indirectly increases the profitability of the game as well.

With PUBG being one of the first adopters of the battle-royale style game, they easily amassed a very large following early on. However, as with most popular game genres, other competing companies wanted to tap into a portion of the market share. Games such as Fortnite, BlackOps 4, and Realm Royale also offer battle-royale modes, attempting to steal some of the user bases of PUBG. With such intense competition, there is little leeway for companies not to create a perfect game. Many companies release games and constantly update them over the years. Studios have to work constantly and closely with their players, to remain in their good graces. If they fail to meet the wants of their player-base, they could easily fall out of popularity.

Problem Statement

Player Unknown's Battleground was released in March 2017 and officially started releasing patches and updating game content in January 2018. They have adopted the strategy of applying a stream of patches and new content to maintain their user base. However, since its release, we found that the user base drop significantly and PUBG has only managed to release 23 content patches in total. To find a possible solution to improve this situation for the developer of PUBG, which is the Bluehole Company. We believe that in these patches, PUBG has either missed the mark on their content or not followed through with the requests of their users. By looking through the reviews for the whole game and for each patch we have chosen, we analyzed the sentiment of users. Moreover, we selected a comparable game to do the comparison. In doing this, we can potentially suggest a strategy for Bluehole developers to regain the bulk of the users they lost in the past year.

Data Sources:

For our data sources, we chose to use Reddit, Steam, and Twitter to crawl information pertaining to our games. Reddit is an online forum that hosts sub-reddits for any category; we

specifically crawled the PUBG sub-reddit for user reactions to individual patches. We performed the same analysis on the PUBG steam forums, to aggregate the overall user sentiment towards individual patches. In order to gather the data from each of the forums, we used Python packages BeautifulSoup as well as Selenium. Steam also collects historical usage data for each of their games such as monthly average users, monthly peak users, as well as the monthly gain and loss percentage. We downloaded this data into excel so that we could put it in a graph to visualize the dataset. For our Twitter analysis, we applied for the Twitter API, which let us crawl tweets related to our game and its competitors. However, we found that the results of Twitter were non-conclusive, so we did not add that in our report.

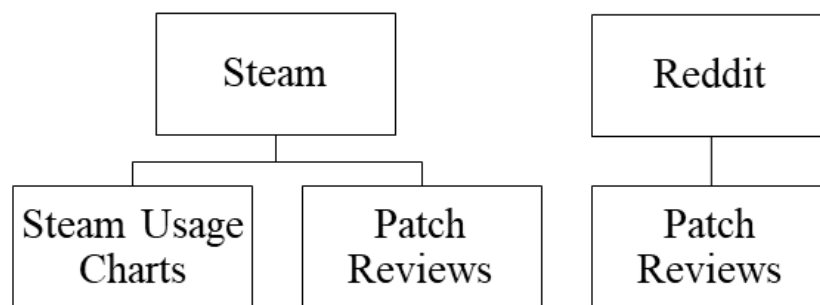


Figure 1: Data Source

Methodology

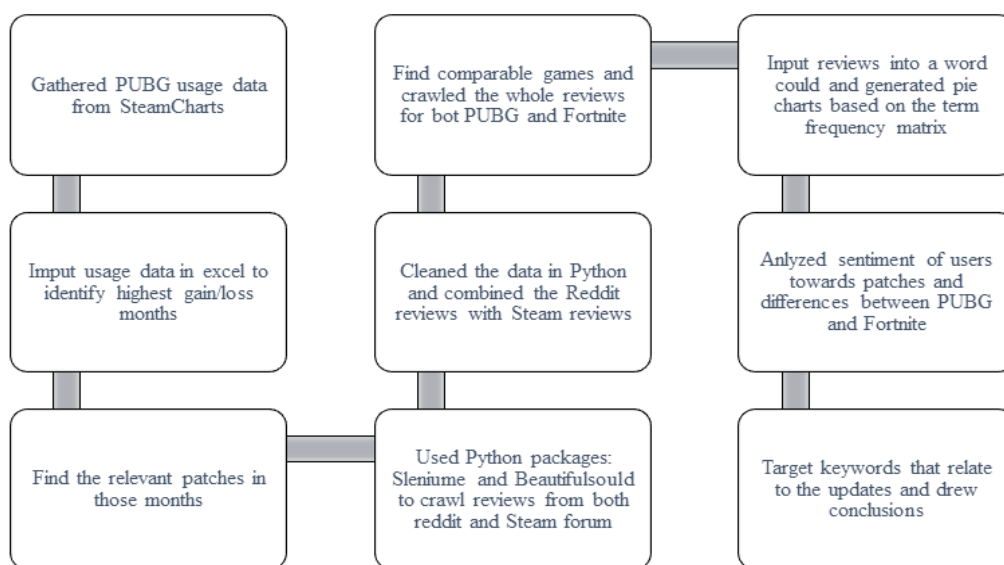


Figure 2: Methodology

Implementation

Steam Usage Analysis and patch selection

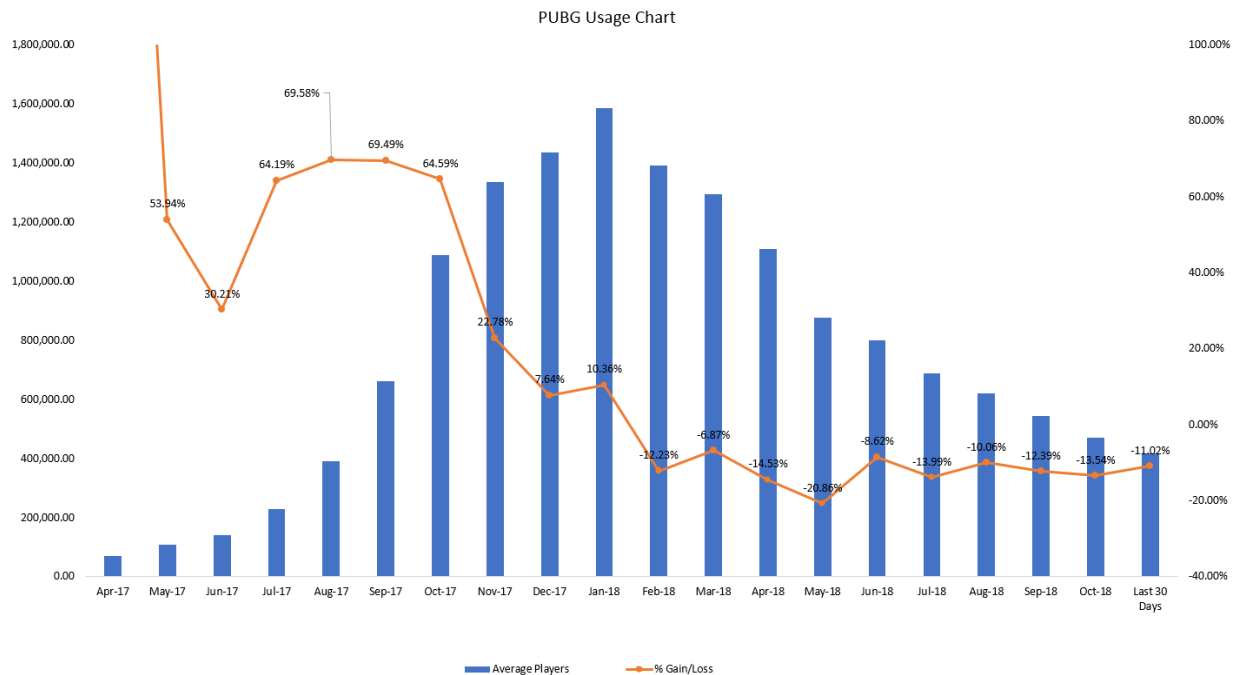


Figure 3: Usage Chart based on Customer Review on Steam

In order to find trends in the usage data, we created a graph of the average number of players per month and created an overlying line graph with the percentage of gain and loss of each month. From there, we identified the key months in which the game had the highest impacting losses and chose to analyze the patches within those months. The game was released to the public as a beta in April 2017. However, the official release of the game was only in December 2017. Patches of the game only began to release in January of 2018. Hence we chose to start our analysis from that date onwards. As you can tell from the graph, PUBG reached its peak number of users in January 2018, when the patches started to release. From there they started a continual decrease in their user base.

We created the percentage line to judge the relative gain/loss of the game, as different months have different numbers of average users it would be unfair to judge the success of each month based on absolute numbers. From these numbers, we identified the key months to focus on, which include:

- January 2018
- February 2018
- April 2018
- May 2018
- July 2018

Crawling method & code

Reddit Code:

Reviews of Each Patch:

There are much fewer reviews for each patch on Reddit comparing with that on Steam. We found the URL for each patch and inspected the reviews to find the HTML parser of the webpages. Then we used BeautifulSoup package to crawl reviews under each URL. We also cleaned the data in Python before it is saved into a txt file. Finally, we saved each review to text file separately.

```
1  from bs4 import BeautifulSoup
2  import urllib.request
3
4  url = 'https://www.reddit.com/r/battleroyalegames/comments/7ufh1h/pubg_news_battlegrounds_patch_notes_update_4_13118/'
5
6  ourUrl=urllib.request.urlopen(url)
7
8  soup=BeautifulSoup(ourUrl,'html.parser')
9
10 print(soup.prettify())
11
12 review=[]
13 for i in soup.find_all('div',{'class': '_3h7_WaV117ooQTH02Uzn2s_s1hmcfrd-0_ckueCN'}):
14     per_review=i.find('p')
15     print(per_review)
16     review.append(per_review)
17
18 len(review)
19
20 New_review=[]
21 for each in review:
22     new_each=str(each)
23     new_each=new_each[28:-4]
24     print (new_each)
25     New_review.append(new_each)
26
27 with open('Review4.txt','a',encoding = 'utf-8') as f:
28     for each in New_review:
29         f.write(each+'\n')
```

Reviews of the game (Fortnite):

```
import praw
from praw.models import MoreComments

reddit = praw.Reddit(client_id = 'Fi4gDecAhoiEgA',
client_secret = 'cr2EjZXSD62Bn-plan2eiUmrICQ',
username = 'tiankongyushang',
password = 'tiankongyu',
user_agent = 'web analytic')

submission = reddit.submission(url='https://www.reddit.com/r/FortNiteBR/comments/a2potr/fortnite_season_7/')

submission.comments.replace_more(limit=None)
# a limit of None means that all MoreComments objects will be replaced until there are none left
for comment in submission.comments.list(): #commentForest
    print(comment.body)

    with open('Reddit Fortnite.txt','a',encoding = 'utf-8') as f: #output, can try Pandas
        for each in comment.body:
            f.write(each)
```

Since the HTML elements of drop down button “More replies” on Reddit are all the same for more replies under each comment and pulling down more replies at the bottom of the page when there is no more comment shown on the current page. We applied the Reddit API to get convenient access to the forum. A Python package called Python Reddit API Wrapper (PRAW) is imported to obtain the script.

Steam Code:

Reviews of the game (PUBG):

We firstly get the URL for the corresponding reviews that are needed. We found that steam requires the users to scroll down on the same web page to view more comments. Thus, Selenium is introduced to do the infinite scrolling in the webdriver by locating XPATH element of “See More Content” and send a Page Down key to scroll down. After all the pages being loaded, the BeautifulSoup package runs to pull data out of HTML pages into text files.

```
import selenium
from bs4 import BeautifulSoup
from urllib.request import urlopen
from selenium import webdriver
from selenium.webdriver.common.keys import Keys
from selenium.webdriver.support.ui import WebDriverWait
from selenium.webdriver.common.by import By
from selenium.webdriver.support import expected_conditions as EC
import time

url = 'https://steamcommunity.com/app/578080/reviews/?p=1&browsefilter=toprated'

options = webdriver.ChromeOptions()
options.add_experimental_option('prefs', {'intl.accept_languages': 'en,en_US'})
browser = webdriver.Chrome('C:/Users/Jstarry/Desktop/chromedriver.exe',chrome_options=options)
#browser = webdriver.Chrome('C:/Users/Lenovo/Desktop/chromedriver.exe',chrome_options=options) #change the webdriver's file loca

browser.get(url)
browser.implicitly_wait(10)

review = []
page = 0

try:
    while page < 10000: #change page
        WebDriverWait(browser,30).until(EC.presence_of_element_located((By.XPATH,"//a[contains(text(),'See More Content')]")))
        scroll = browser.find_element_by_xpath("//a[contains(text(),'See More Content')]")
        scroll.send_keys(Keys.PAGE_DOWN)
        #time.sleep(7)
        print(page)
        page+=1

except selenium.common.exceptions.ElementNotVisibleException:
    print('An error has occurred')
    browser.quit()

finally:
    soup = BeautifulSoup(browser.page_source, 'html.parser')

    for i in soup.find_all('div',{'class':'apphub_UserReviewCardContent'}):
        content =i.find('div',{'class':'apphub_CardTextContent'}).get_text()
        recommendation = i.find('div',{'class':'title'}).get_text()
        review.append(content)
        review.extend(recommendation)

    with open('reviews1.txt','a',encoding = 'utf-8') as f: #output, can try Pandas
        for each in review:
            f.write(each)
```

Reviews of Each Patch:

Different with the infinite scrolling down pattern, users are required to click “next page” button to view the comments on the next page, but the HTML parser of the page stays same as the page goes to next. Due to the number of pages, we still use selenium to run through each page, but BeautifulSoup runs every time as the page is loaded to scrape the webpage.

```
1 # this one is used for crawling comments of each patch
2 from bs4 import BeautifulSoup
3 import urllib.request
4 from selenium import webdriver
5 from selenium.webdriver.common.keys import Keys
6 import time
7
8 url = 'https://steamcommunity.com/games/578080/announcements/detail/1655506336045918090'
9
10 options = webdriver.ChromeOptions()
11 options.add_experimental_option('prefs', {'intl.accept_languages': 'en,en_US'})
12 browser = webdriver.Chrome('C:/Users/3starry/Desktop/chromedriver.exe', chrome_options=options)
13 #browser = webdriver.Chrome('C:/Users/Lenovo/Desktop/chromedriver.exe', chrome_options=options)
14
15 browser.get(url)
16 browser.implicitly_wait(10)
17
18
19 page = 0
20
21 while page < 1472:
22     scroll = browser.find_element_by_xpath("//a[contains(@id,'commentthread_ClanAnnouncement_103582791457492425_1655506336045918090_pagebtn_next') and contains(@class,'pagebtn')]")
23     scroll.click()
24     time.sleep(1)
25     page+=1
26
27     review = []
28     soup = BeautifulSoup(browser.page_source, 'html.parser')
29     for i in soup.find_all('div',{'class':'commentthread_comment_content'}):
30         per_review = i.find('div',{'class':'commentthread_comment_text'}).get_text()
31         review.append(per_review)
32
33
34     with open('Update #5.txt','a',encoding = 'utf-8') as f:
35         for each in review:
36             f.write(each+'\n')
```

User Sentiment Analysis

Analysis for Patches of PUBG


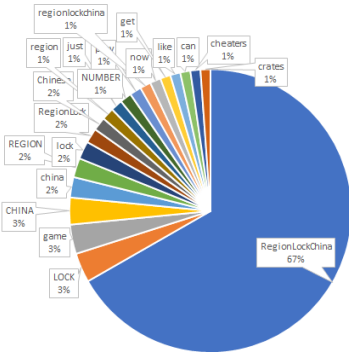
Based on the game perspective, we grouped most frequent words with obvious relation for one same purpose together to find out what kind of things most reviewers concern. We also identified the group of words negative or positive so that we can understand what sentiment they have for each concern.

We decided to analyze the patch #2 in January, since it is the first released patch and also the continuous decreasing of user base started from January. So we assume there are connection between the patch and continuous decreasing of user base. And we want to verify our assumption by doing sentiment analysis of reviews for patch #2. We also decided to analyze all 3 patches in May, because there was a big percentage loss of user base in May. We assume the 3 patches have relation to the loss, so we want to verify our assumption by doing sentiment analysis of reviews for patch #12, #13 and #14.

Patch #2 in January

With the game background, “lock” is obviously a negative word in a context of PUBG.

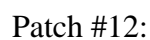
Category	TF rate
RegionLockChina LOCK RegionLock regionlockchina RegionLockCHINA	22037/31038

Though patch #2 added a new way to report suspected cheaters in game, according to the reviews, the updates of patch #2 did not help to stop the cheaters in the game.

Patches in May

Item	Percentage
MICRO	13%
TRANSACTIONS MORE	13%
game	9%
FIX	6%
ADD	5%
game	5%
TRADE	4%
LET	4%
game	4%
skins	4%
ITEMS	4%
DUO	4%
M16A4	3%
RECOIL	3%
PLEASEFIX	3%
UMP	3%
NERF	3%
update	3%
THANOS	3%
fix	3%



“BLURRY” “SCOPES” is a phrase with the continuous word in the game, what it means is that the scope for the gun in the game is not clear so that they cannot experience well in the

game. “Fix” shows that since there are imperfection existing, people ask to fix the imperfection. So “fix” is a negative word in the game. “Nerf” is one of the fixing bug method, which is a negative word too. Since it is a root form and normally it means people ask the company to nerf something in the game.



Patch #12 made a significant changing with weapon balance. But people comment about Blurry scopes and ask for nerfing powerful things at 122/5946 TF. Which implies that for some weapon, such as AKM appearing at 15/5946 TF, it is still not balanced for users to use.

“FIX,” “PLEASEFIX,” “need” indicated that there is imperfection existing, people ask to fix the imperfection. Consequently, these words are negative words in the game.



Category	TF rate
MICRO TRANSACTIONSMORE	770/11619
FIX PLEASEFIX need	342/11619

Based on above, Patch 14# fixed nine bugs for the game, but words such fix please fix and need still appear in the reviews at 342/11619 TF. It implies a certain level of users; they still need the games to fix some bugs or do some change. Also, “MICRO TRANSACTIONSMORE” is not related to the content updated in the patch #14, but since people even come to ask it back under patch update page and the phrase appeared here at 770/11619. It implies that it could be a related factor that affecting the reducing of the user base.

“Nerf” is one of the fixing bug method, which is a negative word. Since it is a root form and normally it means people ask the company to nerf something in the game.

“Good,” “better” are straight words representing a positive sentiment. They implied that some people think this update helped them improve their game experience.

Category	TF rate
NERF	181/5907
Good better	49/5907



Overall, Patch #14 fixed the sound of footstep and kept adjusting weapon balance too. But there are people still asking fixing more.

Comparison with Fortnite

To complete the analysis, we found another game, Fortnite, which is related to the PUBG’s user base’s change. Fortnite as another battle-royale game released almost at the same time as PUBG, while PUBG showed a drop in user base, Fortnite gained. Fortnite and PUBG have a ton in common. They are both multiplayer online shooters that allow for solo or team play and subscribe to the battle-royale format, which drops 100 players at a time into an enclosed space where they must scavenge for weapons, medical supplies, and other resources to be the last one standing. However, when we focus on the differences between them, the differences may imply some important factors to the user drop of PUBG.

Firstly, the two games have a different user base, more than half of the players are Chinese for PUBG, while players equally separated in France, Denmark, Sweden, etc. This difference in players’ location may be an implication of the cheaters in China drove out many foreign users in PUBG, which we found in the previous analysis.

Secondly, Fortnite has more interesting playing modes than PUBG, which contains only pure competition. Players in Fortnite can build their base and traps to defense enemy in a more free and customized way. Moreover, every two weeks or so, a new temporary game mode becomes available to play. A few examples include 50v50, five teams of 20, a snipers-only mode, an explosives-only mode, and Blitz, which speeds up resource gathering and the speed of the storm. Temporary modes tend to be more experimental, while, PUBG only have several modes in total. That’s what the developers in PUBG should improve.

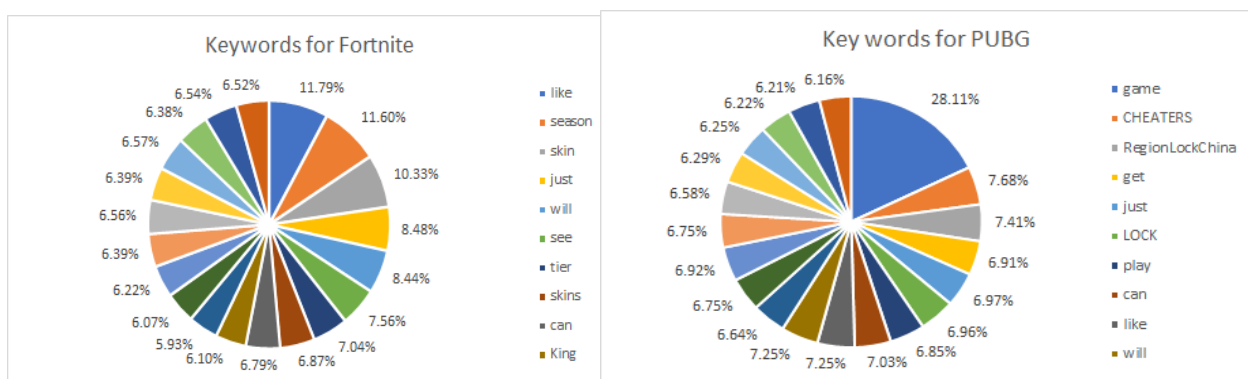
The last but not the least, Fortnite has 1,180 patches. However, PUBG only has 23 for now. With more patches, the company can have a quicker response for users’ complaints, bugs,

and even create more events or modes. The frequency of releasing the patches means the developer's ability in maintaining the games and the endurance of the game. The developers for Fortnite is Epic Games, which is a much bigger company than Bluehole. That's maybe an explanation of the weakness for PUBG.

For the next step to further our analysis in comparing these two games, we looked through the reviews in Reddit for both games and generated into the wordcloud. In general, we've got many positive words in the wordcloud for Fortnite in contrast to the negative ones for PUBG, such as 'fix,' 'fps,' 'try,' etc. mean the game needs more improvement. Look through the detail in wordcloud for Fortnite, 'skin' means Fortnite has a lot of skin choices for users to custom themselves. This one can combine the change that Bluehole deleted the micro trading transactions even to restrict users. Moreover, Fortnite has a seasonal changed map selection while PUBG only has three maps in total. While looking through the actual review for Fortnite with the word "season", we found that users have a positive response to the seasonal changed map. In contrast, PUBG even closed the map selection in patch 17#, which the developer for PUBG should pay more attention in its users' demand.

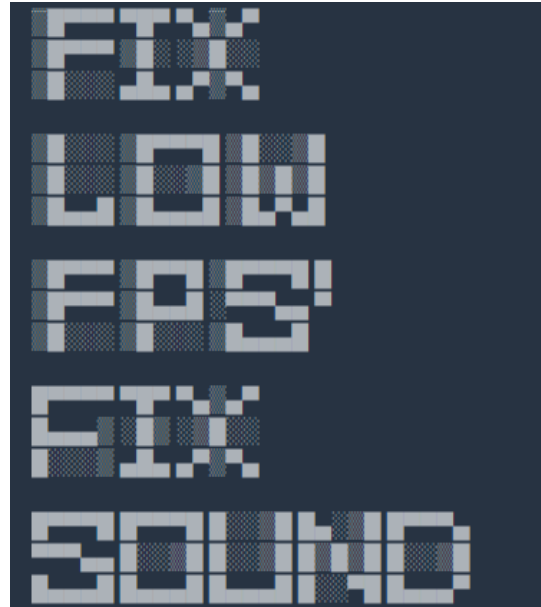


The term frequency matrix reflects the overall response of the comments for two games. In PUBG, despite the ‘game,’ ‘CHEATERS’ and ‘RegionLockChina’ have the highest term frequency, which reveals these cheaters in China is always the most important factor that the users care, maybe the biggest factor that drove so many users out. With cheaters in a game, the one will have a superpower such as perspective, automatic aiming, through-wall, etc., but the others will have a terrible game experience, which may lead to them abandon this game. In fact, Chinese are professional in inventing and selling hackers, that’s why so many users mention ‘Region Lock China’ in the reviews. The biggest problem users responded it should always be the most severe problem that the developer need put in first place.



Things to note

Many of the comments we found on the steam review site were written in symbols and foreign characters. Based on the frequency and size of these comments, it is obvious that the users were passionate about getting their message heard. However, due to the words being symbols, we were unable to read the messages with our text mining code. One suggestion for the users would be to rewrite their comments for them to be picked up by the text mining code. On the other hand, we could also find another package in order to be able to read these messages.



Conclusion & Recommendation

Through our analysis, 'cheaters' is the biggest cause that drives other users away from the game. It causes people with negative feeling to the game and could be related to the decreasing of user base. In the patch 2#, the first officially released patch, where users can firstly post their expressions to the developer, most of the comments mentioned this word as well. This condition verifies the result that 'cheater' and 'China' are the factors that the developer should find a way to remit the conflict. Build another server only for Chinese users as what the comments said is a way that the developer should think. Although it may take a risk while doing so, since we do not know how the majority of people in the game are distributed. Such risk may be our future direction to find the distribution of users. More importantly, prohibiting from the hackers is another suggestion that the company should work through later. Although Bluehole has taken actions in preventing cheaters from the game, what they have done is far more behind the improvement of those advanced hackers. Find another way to strictly prohibit the hackers is the key for PUBG to regain its losing players.

Creating more talented modes is always a good way to attract users as what Fortnite has done. Providing more selectable maps and alterable appearance are good directions to make the game more attractive. A mode which is freer and more available for users to explore can resistant to users' psychology as fickle in affection. Being innovative all the time is the only way of survival for a game to remain popular in today's intensive game market.

Furthermore, patches need to be released more frequently and more relevant to users' needs. From the analysis in patches, users expressed their perspective in reviews for what they do not like and wanted the developer to improve, especially bugs in the game. Since bugs cause the bad experience of users, the company should keep working on fixing bugs. Tracking the bugs is the neediest task by analyzing the reviews for the new patch. Moreover, the customers are always the key to a business's success. Matching what the users need is the only way to maintain the user base. However, the reviews of some patches revealed what the patch did was going backward from people's expectation, which then caused disappointment and deviation. The developer could establish a group of people collecting reviews from users and summarizing what they should improve.

Future Direction

According to the sentiment analysis we have done so far, we noticed that Chinese users and the shutdown of Micro-transaction could also be two other factors having a significant influence on users' base of PUBG. Consequently, we plan to do more research on these two factors and find out if there is any relationship between them and the user base.

Since there are so many images reviews that we cannot crawl as text but also express users' sentiment in the reviews page of each patch, we plan to find a way to gather all the images information from the website to do more accurate sentiment analysis.

Recently, PUBG's patch 24# is planning to release a new map, Snow, and it also fixes some bugs and create new playing modes. The map is now available in the test server. We can analyze the usage after the patch is released and based on that, to prove further our conclusion that more map selections and more modes will regain its users.

References:

1. Reference for number of players monthly changes
<https://steamcharts.com/app/578080#6m>
2. Reference for usage chart

https://store.steampowered.com/app/578080/PLAYERUNKNOWN'S_BATTLEGROUNDS/

3. Fixing bugs: A bug is a logical error in the programming of a computer program. A bug has nothing to do with a virus which is completely different and computer viruses themselves can have bugs because they are also a computer program. An example of a bug in a game is something like being able to walk through the wall or under the floor. Fixing bugs means to improve the programming and make everything smoothly in the game.
4. Cheaters: It is a program that designed to have some super power in the games and violates the balance of the games.
5. Region Block China: Basically, Chinese is professional in designing and selling cheaters, so when you meet a person who have cheaters in the game, it is more likely to say that person is a Chinese. That can be explained, when people say region lock china, it means lock a server for Chinese, they don't want to be bothered by the cheaters.