DOTA2 SENTIMENT-USER ANALYSIS

Evan DeCastros, Chien Min Wang,
Johnathan Jackson, Yidan Hu, Zhen Wu

DOTA2 Sentiment-User Analysis

Executive Summary

Video game companies face a highly competitive landscape, where game designers have a few options for maintaining relevance in the market. Designers can either create new games every few years to pull in revenue and draw customers back to their products, or they can continuously update and improve their game over a longer period of time. One game that fits the latter business template is DOTA2, created by the gaming company Valve.

DOTA2 is an ongoing multiplayer game that was created and launched in 2012-13. Since then, it has seen continuous updates (usually at least one per month) and internationally sponsored tournaments in attempts to keep users interested. Using data provided by Steam - a platform on which the game is played - we can see that the average monthly users playing DOTA2 saw a steady increase from its inception to its peak in February 2016 when it hit a high in average monthly users of over 700,000. However, since then the number of average monthly users has been steadily declining, to where it now stands at just under 500,000.

Our goal in this project was to analyze the usage trends of this widely popular game and try to find correlations between those trends and the sentiments of those users. Using pattern analysis, we hoped to find the months that saw the highest fluctuations - either positive or negative - in average monthly users. Using these months, we cross-referenced publicly provided timelines of updates and tournaments to find specific key dates to analyze, for which we performed sentiment analysis of the user comment forums on Steam. By finding correlations between user traffic and the sentiments of users in reaction to game updates and/or tournaments, we can be able to identify the most effective means of drawing users back to DOTA2 and keeping them playing, thereby increasing revenue and satisfying the business needs of Valve.

Through our analysis, we found that small updates to gameplay and large-scale tournaments were the most effective means of increasing average monthly users and the sentiment among those users. Conversely, the effects of substantial game updates and smaller tournaments were muted, if not inverted. As a result, our recommendation to Valve would be to pool their efforts into creating more large-scale international tournaments throughout the year. Additionally, we would advise focusing their efforts on this rather than on radical changes to gameplay, while keeping their usual monthly-bi monthly minor game update schedule intact. We believe that this strategy will result in the continued success of DOTA2, as well as the optimal enjoyment of the game by its users.

DOTA2 Sentiment-User Analysis

Introduction

Increasing access to technology has been a boon to people across the globe. As more and more people around the world gain consistent access to technology, the companies that provide that technology have been some of the biggest winners. However, the same increased access and demand for technology has also created very competitive markets. One industry especially affected by this shift in competition is the video game industry. Tasked with navigating both the entertainment industry and the technology sector, companies in the video game industry face very unique challenges in remaining competitive. The few video game companies that are successful enough to generate sales that make up for their development costs must immediately think about what they will do next. With a constant stream of games being released during a given year, companies must always have a strategy to retain customers. Companies that fail to find an effective strategy will find themselves irrelevant, and unable to generate profit in a landscape filled with aggressive and effective competitors.

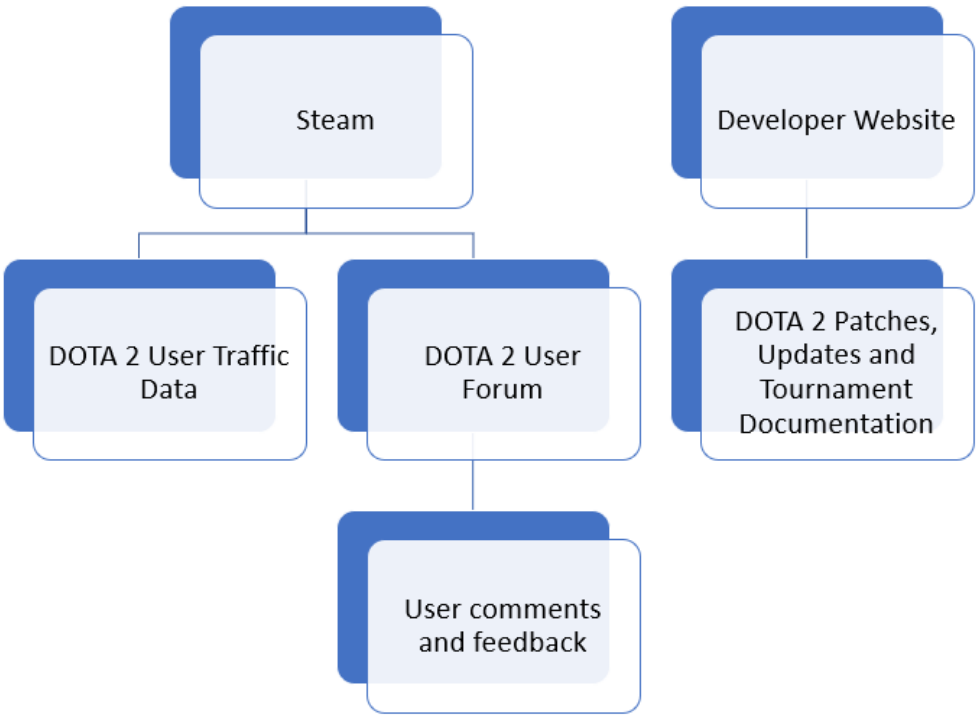
Problem Statement

Most video game developers have two basic strategies: Develop a franchise and release new, modified iterations of the franchise game over several years, or create a massive base game and release modifications or add-ons to the original game to keep up with changing times. DOTA2 is an online multiplayer role-playing game that fits the latter strategy. DOTA2 was developed by Valve Corporation and officially released internationally in July 2013. DOTA2 has a fairly simple premise that led to its rapid increase in popularity. Games of DOTA2 consist of two teams of five players each defending their own positions on a map. Players get to control a “hero” character with unique skills and abilities, and use that hero to form a strategy to defeat the other team.

Unique and fun characters and inviting gameplay made DOTA2 popular in its early years. However, DOTA2 followed the latter strategy that was previously mentioned, and as the years have gone by the game has seen a decline in user traffic. Since Valve decided to release DOTA2 as a dedicated platform, they have relied on patches and updates to gameplay to keep the content fresh and appealing. Financially, this puts Valve in a very restrictive predicament. The DOTA2 base game is completely free, and Valve only makes money through microtransactions (for cosmetic items that do not greatly affect gameplay) and tournament sponsorship. Because only a subset of users choose to make any of these purchases, Valve is heavily reliant on a substantive inflow of users who will be tempted to buy new cosmetic items. Intuitively, if Valve cannot maintain enough DOTA2 users, they will not make as much money. Valve depends on these buyers to subsidize the costs of running servers and developing content for the entire DOTA2 population.

Today, DOTA2 is over 4 years old. Valve must address the issues that are costing them user traffic, and locate the strengths that kept DOTA2 popular in the early years. We can use web analytics techniques to maximize Valve's efforts to keep DOTA2 alive, and hopefully regain the user traffic that it once was able to boast. Steam, the platform that hosts DOTA2, keeps detailed record of DOTA2's historical traffic and the comments users have on the game. Through combining this information with the update information supplied by Valve, we can find out what is making users happy or upset with the game. Then, we can address those issues to both retain and attract users, in order to bring Valve future revenue.

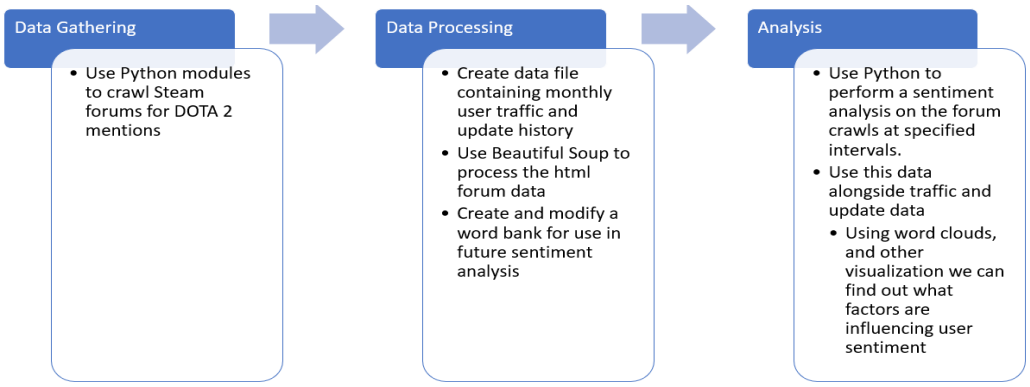
Data Sources and Description



Useful data analysis is always dependent on robust and mature data. Fortunately for Valve, this data exists for DOTA2. Transaction level data for DOTA2 is unfortunately proprietary and we were unable to access it. In place of that data, we are operating under the business assumption that the number of users is positively correlated with the number of transactions. The number of concurrent users per month is available directly from Steam, and we were able to directly record this numerical data to excel. Using this data, we were able to calculate relative and absolute monthly change in user traffic. For each month of user traffic data, we visited the developer website to find data for any and every patch or game change that happened during that month.

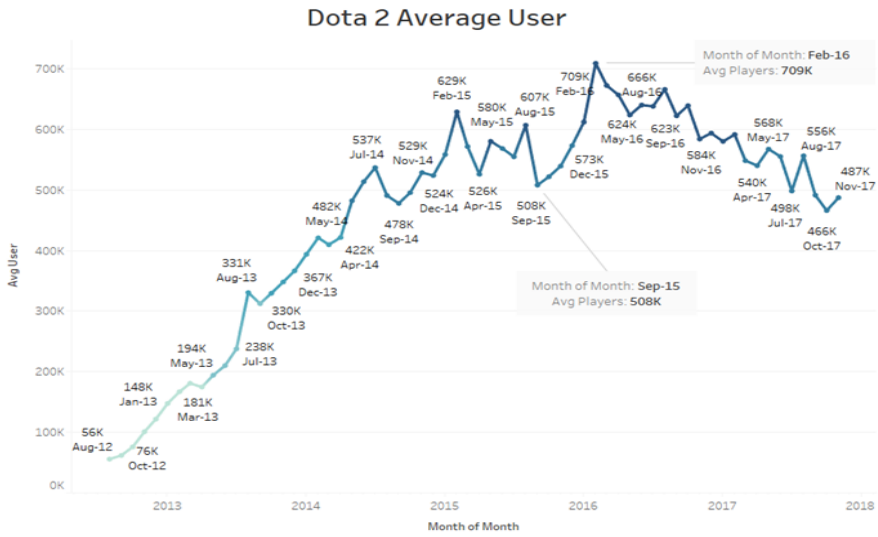
Most importantly, we used Python to crawl the DOTA2 forums on Steam for user comments about the games during months with updates. We ended up with a csv file containing string data for each comment in a given time period. That data was used later to gain an insight on the feelings and comments of users. It is that data, combined with our knowledge of the user traffic that allowed us to make relevant recommendations to Valve for user retention.

Methodology



Steam User Analysis

For the purpose of finding trends in the user patterns of DOTA2 over the last few years, we can use data provided publicly by Steam's sister website steamcharts.com. Steamcharts provides the average monthly users for any game, along with the peak users for each month as well. Below is a graph we plotted using this data to show the average number of users on the y-axis with a timeline on the x-axis to see how the user activity has changed over the course of DOTA2's lifecycle.

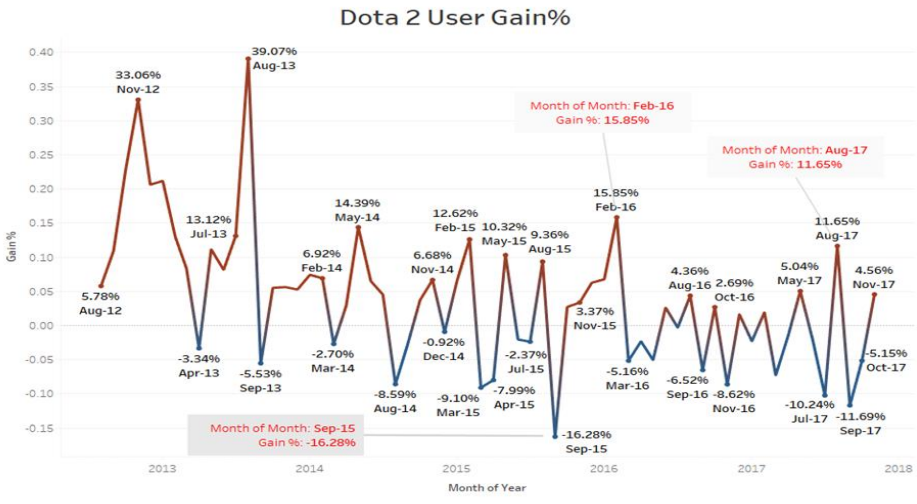


As can be seen above, the game was created and released (in a beta stage) in August of 2012, with initial average monthly users of 56,000. In August of 2013, the beta stage was officially closed and the game was released to the public, resulting in huge increases in average monthly users as the game became available to a larger audience. For the purposes of this analysis, we will begin our search for inflection points in gameplay after this time, as it is obvious why the game users increased so rapidly during this time. Additionally, this is not a repeatable process - the game can only be released once and this increase is therefore a onetime event, unusable by Valve in the future.

In order to hone in on specific months as targets, we created additional attributes for the data provided by Steamcharts called “Gain %” and “Absolute Change”. Gain % is calculated as the percent change in average monthly users from the previous month to the current one, and Absolute Change is the absolute percentage change in average monthly users. For example, if the average monthly users in the previous month was 100,000, and the following month it was 75,000, the Gain % would be -25% and the Absolute Change would be 25%.

Using these new attributes, we can find the months that saw the largest changes in users playing DOTA2, and then search those months for key events (such as game updates or tournaments) that may have been a cause of the change. We decided to set a threshold of a 10% absolute change in order to investigate a month. While the number is somewhat arbitrary, we believe that 10% is a significant trend change in the average number of users playing a game. Additionally, this criterion neatly accounts for 8 out of the 51 post-beta stage months that DOTA2 has existed for, or about 15% of the months. This narrows down the lifecycle of the game enough for us to find some key dates, while also providing a long enough timeline to search through and accurately capture different ranges of sentiments.

Below is a similar graph to the average monthly users as previously shown, but focusing on gain % instead:



Here you can see the months that have the most extreme changes in users playing DOTA2. Using the months with the largest changes, we hoped to find some inflection points, which would allow us to focus on specific dates for our sentiment analysis. Examples of a few months that fit our criteria are:

- May 2014
- May 2015
- September 2015
- February 2016
- August 2017

Identifying Key Dates

After recognizing certain months with significant user fluctuations, we referenced the public timeline of updates and tournaments to find whether there were updates and/or events that might have caused the changes in these months.

Overall, seven periods were selected for further analysis. According to our references, big changes in May-2014, May-2015 and Sep-2015 may be related to updates, and changes in Feb-2016 and Aug-2017 may be connected to premier tournaments, which are international tournaments of a large scale such as The International and the DOTA Major (especially Shanghai Major). In addition, we added another two periods that needed to be analyzed, although the user gain in these periods did not fit our criteria. The first is Dec-2016, when one of the biggest updates of DOTA2 was released which changed nearly every aspect of the game. Unfortunately, the user gain in this month was only 1.65%, so we want to find whether the users' feedback could give some insight into its reception. Nov-2017 is another point we put our attention on because it's the latest data we have, and could be related to the most recent large-scale update which took place on 10/31/2017.

For these events on specific key dates or during certain periods, the next step is to figure out whether these events have caused large changes in DOTA2 user activity. We chose to crawl comments from the DOTA2 forum websites where official news about events are released each time, as users can write comments below to express their feedback. After the crawling process, sentiment analysis can be done for each cloud of website comments.

Below is a table of our findings which correspond to the seven periods mentioned, along with the event we believe is connected and the website page with the comments that were crawled:

Month	Gain%	Event	Website

May-14	14.39%	Update on 4/29/2014	http://steamcommunity.com/games/dota2/announcements/detail/1733067862481822920
May-15	10.32%	Update on 4/30/2015	http://steamcommunity.com/games/dota2/announcements/detail/240158195517931548
Sep-15	-16.28%	"Reborn" Update in September	http://steamcommunity.com/games/dota2/announcements/detail/800867765737127072
Feb-16	15.85%	Shanghai Major	http://steamcommunity.com/games/dota2/announcements/detail/843667822464962362
Dec-16	1.65%	Biggest update ever in 12/12/2016	http://steamcommunity.com/games/dota2/announcements/detail/528441854359282755
Aug-17	11.65%	The International 7	http://steamcommunity.com/games/dota2/announcements/detail/1440443348072170137
Nov-17	4.56%	Update on 10/31/2017	http://steamcommunity.com/games/dota2/announcements/detail/1449457773770927103

Steam Forum Crawling Method

Physical system design: Using python to build a web crawler to scrape the comments from Steam.

Conceptual design:

1. Inspected the requests the page made through the developer panel, and found the URL which contains unpacked object of Json.
2. Imported package of python, Json, that the data can be loaded as dictionary type, and extracted all comments from Json type.
3. Filtered the non-English terms by using package, Enchant, so that the result can be used to do lexical text mining.
4. Used package, TextBlob, to do sentiments analysis

Part of code:

```
comment_list=[]

def is_english_word(word):
    dictionary = dict.fromkeys(nltk_words.wrods(),None)
    try:
        x = dictionary[word]
        return True
    except KeyError:
        return False

for i in range(00, 970, 10):

    url = 'https://steamcommunity.com/comment/ClanAnnouncement/render/103582791433224455/1449457773770927103/?start='+str(i)
    r = requests.get(url)
    data = r.json()
    soup = BeautifulSoup(data['comments_html'], 'html.parser', from_encoding = 'big5')

    words = set(nltk.corpus.words.words())
    english_words = []
    d = enchant.Dict("en_US")
    try:
        for div in soup.find_all('div', {'class': "commentthread_comment_content"}):
            string = str(div.find('div', {'class': 'commentthread_comment_text'}).get_text().strip())

            for word in string.split():
                if d.check(word):
                    english_words.append(word)

                comment_list = ' '.join(english_words)
        print(comment_list)

    except UnicodeEncodeError:
        pass

testimonial = TextBlob(comment_list)

print('Result')
print(testimonial.sentiment)
print("Polarity of all Comments:" +str(testimonial.sentiment.polarity))
```

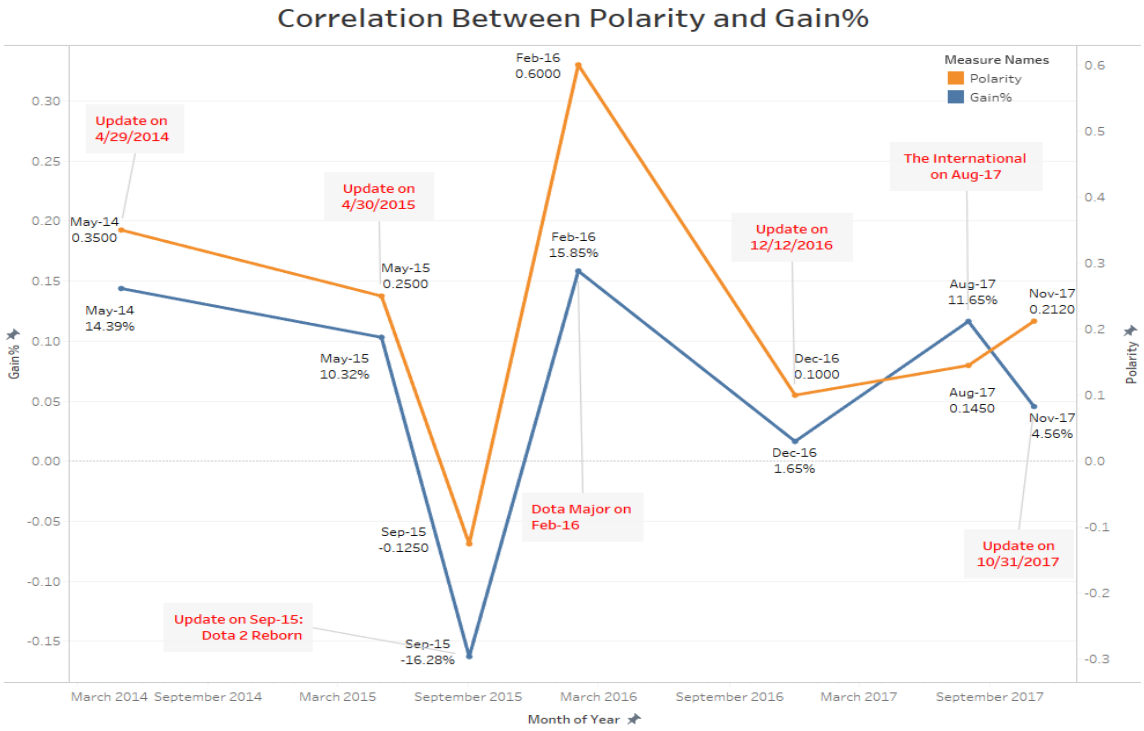
Output: (Included comments and result of sentiment analysis)

```
why complaining stop playing 9 1.8
Rip Rip jungle. Nice hope I get to play it soon How big is the check it out at 1
2 but I didn't get it 3 1.3 welcome to
Bane has a new talent that allows Enfeeble to steal enemy damage instead of just
 reduce it. TRASH Bane doesn't steal he reduces the damage. rip ice frog kill ic
e frog 2 BANE OP 2 with his HIS AFTER CARRY ITEM BUILD steal bog good
So how much this is the worst update so far. what thew did do to the are you an
that must make changes just so someone can keep their job and say they were the
bottom may as well be placed on top of one another. whoever this crap hard. Dark
 Willow 2 copy Paladins IAN ice frog this update so fucking cancer and is EVERYW
HERE
rip update wait for me coming back is getting now HOW CAN I FARM LC WITHOUT MY B
EST
cancer league is better High Seas is that yeah 1 intensifies. Going to era... no
 battle pass this Time to purge like every update in the past 2 years
invoker's moving you will miss the battle pass Cancer update. 7.07 Update GOOD T
hanks for the update. U WITH UR FAST PACED Why remove VALVE why remove
I want Half Life. Nice new hero come up with immortal right when it release wow
GIVE ME Why is 4 is 3 Q U I C K M A T H ITS HERE rampage is this 4
Just in time WHY IS EVERY NEW HERO SO OP VALVE STOP TRYING TO SELL YOUR DUMB BUN
DLES SO FAST Go time boys
Just in time WHY IS EVERY NEW HERO SO OP VALVE STOP TRYING TO SELL YOUR DUMB BUN
DLES SO FAST Go time boys
Just in time WHY IS EVERY NEW HERO SO OP VALVE STOP TRYING TO SELL YOUR DUMB BUN
DLES SO FAST Go time boys
Result
Sentiment(polarity=-0.012878787878787878, subjectivity=0.5181818181818182)
Polarity of all Comments:-0.012878787878787878
```

Analysis of User-Sentiment

Overall Sentiment Analysis

Having retrieved sentiment analysis results from user comments, we used "polarity" to measure the users' sentiments: Polarity has value between -1 to 1; the sentiment is positive for polarity larger than 0, negative for polarity less than 0, and the closer polarity is to -1 or 1 the stronger the sentiment is. We then produced a line chart to show the overall relationship between sentiment polarity and user gain. As shown below, the yellow line represents sentiment polarity values between -1 to 1 and the blue line shows user Gain % in corresponding months with a timeline on the x-axis. Specific values and annotations of events are marked:



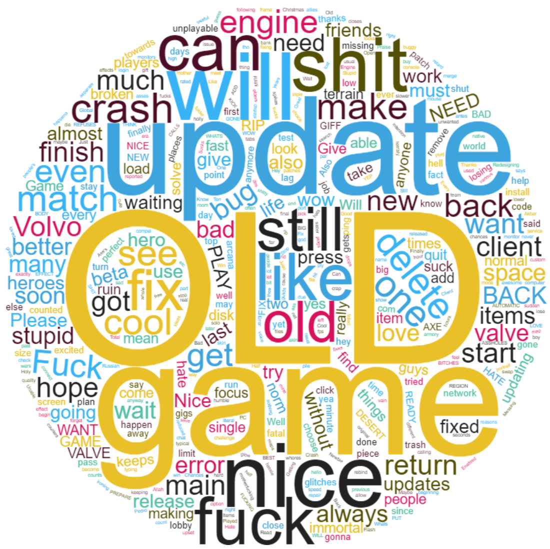
Overall, the polarity and user gain show a strong correlation. There existed large increases in users when the sentiment analysis showed that users were satisfied about updates/tournaments at those periods and vice versa, for example in September 2015 and in February 2016. If users only had indifferent comments for events, the user gain also experienced mild changes, such as in December 2016.

For further evaluation we did more analysis of comment sentiments for both game updates and tournaments individually, and visualized our analysis results.

DOTA2 Game Update Analysis

We attempted to figure out what kind of result some significant updates bring to this game. One of the biggest updates was released on Sep 15, 2015, called "the Reborn". This update brought a brand-new engine to DOTA2. We know the game is

based on an “engine” to run, and thus, everything that players were familiar with was changed such as HUD and game performance. Below is a word cloud of the comments from the analysis, drawn on wordclouds.com:



We can clearly see that “OLD” is big in the middle. And we also see “hope”, “back”, “return” in relatively obvious places with decently large font. We interpreted this as the players “hope for their old game engine back.” With NLTK sentiment polarity -0.125, the players’ attitude toward this big update is negative. We can also see quite a few impolite words in this cloud as well.

Nov-15	539,536.30	17,594.60	3.37%	943,635	3.37%
Oct-15	521,941.70	13,784.90	2.71%	917,306	2.71%
Sep-15	508,156.80	-98,787.10	-16.28%	888,728	16.28%
Aug-15	606,944.00	51,953.00	9.36%	933,942	9.36%
Jul-15	554,991.00	-13,457.30	-2.37%	877,264	2.37%

The user traffic during this month shown above is consistent with our sentiment analysis. With this screenshot from Excel we can see DOTA2 suffered from a 16.28% decrease in user traffic this month, even though there was only a 9.36% increase the previous month due to The International Tournament which we will mention later. This means the decrease was almost double as big as the increase from the tournament. It is not hard to imagine players stopped playing DOTA2 at least temporarily after this big update, and maybe some old devices were not able to run the new engine perfectly. We believe too large-scale changes such as changing the platform in this way is not a good idea to attract and keep users.

Another big update was released on Dec 16, 2016, but this one was more about the game content instead of the platform. The sentiment polarity of this patch was 0.1.

Though there was a correlated increase in user traffic, the increase was not as big as we expected.

Feb-17	591,567.30	11,285.80	1.94%	1,040,877	1.94%
Jan-17	580,281.50	-13,639.10	-2.30%	1,007,451	2.30%
Dec-16	593,920.60	9,669.30	1.65%	1,014,671	1.65%
Nov-16	584,251.30	-55,103.80	-8.62%	1,007,270	8.62%
Oct-16	639,355.10	16,771.20	2.69%	1,141,191	2.69%

The word cloud below can help to answer our confusion. Increasing players can also bring new problems such as lagging. You would never want the hero that you are operating to take a full second to react to your command. Internet and server problems are something the game company ought to deal with, especially after an update is released. Even though we have a positive increase, there are still some impolite words in the cloud. And we can see players using “RIP” to mock that this game is dead. However, not every player can be satisfied with an update, since everyone has their own opinion about the game. Maybe the heroes they loved to choose are “nerfed” in consideration of game balance, or maybe the heroes they dislike are “buffed”. Regardless, there was a positive reaction to this update in terms of both sentiment and user traffic, even if it was a smaller reaction compared to some other updates.

DOTA2 Tournament Analysis

Tournaments are the biggest events of DOTA2. Pros play hard for high bonuses, and players cheer for their favorite clan. These events can ignite the player's passion for the game and bring more heavy traffic. In Feb 2016, the Shanghai Major was held in Shanghai. Even though the Chinese teams played poorly at the tournament, we still found a sentiment polarity at positive 0.6. As supported by our sentiment analysis, our word cloud shows that user feedback was generally positive, with “nice” being the most frequent word in the forum. From this, we can assume that hosting tournaments is a great way to retain positive sentiment.

At last, we want to mention The International tournament, the most important event in the world of DOTA2, the feast held in Seattle every August. The result of our analysis supports the correlation shown in the Shanghai Major - sentiment was also positive during The International. From the table below, we can see the tournament was able to create an 11% gain in users following a disappointing 10% loss in the previous month, although user traffic the following month decreased by 11% as well. The International pulls a great number of users back to game even if their stay is only temporary.

Oct-17	466,126.80	-25,323.40	-5.15%	832,550	5.15%
Sep-17	491,450.20	-65,046.10	-11.69%	829,555	11.69%
Aug-17	556,496.30	58,051.90	11.65%	876,395	11.65%
Jul-17	498,444.40	-56,844.70	-10.24%	824,297	10.24%
Jun-17	555,289.10	-12,237.40	-2.16%	923,122	2.16%

Positive words in comments had a much greater frequency than in the previous tournament as well:

Conclusions

From our analysis, there were severable notable conclusions that we were able to draw:

- Large-scale tournaments lead to an increase in both average and peak monthly users, while also generating positive sentiment among users
- Smaller tournaments and in-game events have a much smaller, if not negligible effect on average monthly users
- The positive effects of tournaments are temporary
- Small gameplay updates generate positive sentiment and can also increase average monthly users
- Large-scale updates such as game engine overhaul or platform changes generate negative sentiment, and can lead to temporary declines in average users

The effects of tournaments are somewhat obvious - that massive international tournaments have positive effects on both sentiment and users. This was expected, as the tournaments provide incentive for more users to get involved - whether through watching the gameplay unfold or participating in it themselves. What we were not expecting regarding tournaments was for smaller ones to have such a minute effect. There are dozens of smaller tournaments across the world each year, but the only ones that we found to have the positive impact that we were looking for were The International and The Shanghai Major. These are two of the largest - if not largest overall - DOTA2 related tournaments.

If smaller tournaments do not provide bumps in users playing DOTA2 or the sentiment regarding it, then it does not make much sense for these tournaments to occur at all. Our recommendation based on this evidence is for Valve to try and cut down the overall number of tournaments that take place throughout the year, and build up a few larger-scale ones. If Valve can have 4-5 yearly tournaments close to or at the level of The International, they can be able to generate what would amount to a quarterly tournament bump, where every few months there is a massive tournament that brings DOTA2 back into the spotlight. Currently this only occurs a couple times year, so doubling the frequency of these can have a massively positive impact on the number of users playing the game, which can indirectly lead to significant increases in revenue for Valve.

Compared to tournaments, our analysis indicates that the effect of updates is the opposite. Smaller game updates have a more positive effect while larger game updates have a more negative effect. With smaller updates, they are likely focused on bugs and small enhancements to gameplay that users have been asking for. These small updates are easy to implement, and can go a long way in improving the user experience. The negative effects of larger updates seem logical as well. If there is a massive update to the game that drastically alters the gaming experience, common sense tells us that there are likely to be gamers that are annoyed by this. With a game that has been played for almost a half-decade, users become accustomed to a specific gaming

experience. When that experience is somewhat radically changed, a negative sentiment is to be expected. Further confirming our analysis is the fact that the negative effect on users from this is temporary according to the Steam user logs. It seems that though users may avoid the game for a short period of time after a radical update, eventually they come back and get used to the new gameplay, with the average monthly users perking back up shortly after.

With these effects of game updates now known, we would recommend that Valve keep their overall number of updates about static, but cut down on the major updates as much as possible. If Valve is able to only release major gameplay updates once or twice a year, they can cut down on the negative backlash that generally follows and keep their users happy and playing. As for the smaller updates, there are already at least one a month usually, so we would recommend that Valve keeps releasing these as they have been.

We believe that if Valve were to follow these recommendations, they would see a hugely positive effect on their number of users who play DOTA2, as well as on the sentiments of those users regarding the game. Having more large-scale tournaments throughout the year can help in keeping users playing as well as attracting new users, while maximizing the number of small game updates and minimizing the number of larger ones can keep those users happy and thinking about DOTA2 positively. For a game that has been around for years and faces new competition constantly, these are necessary developments for the survival of the game.

References

Timeline of DOTA2 updates and tournaments. Retrieved from:
http://wiki.teamliquid.net/dota2/Patches
http://wiki.teamliquid.net/dota2/Portal:Tournaments

DOTA2 Steamcharts user traffic data. Retrieved from:
http://steamcharts.com/app/570