Analysis of The Best Year for Club Penguin Parties

Parker Dingman

12/16/2024

Introduction

Club Penguin was an online multiplayer game that was available from October 2005 to March 2017. The game was targeted to children aged 6 to 14, although millions of players across all ages played as well. The game involved visiting different rooms, areas, mini-games, and chatting with other online players with a controlled, kid-friendly chat feature. In addition, players could personalize their experience by creating a custom penguin avatar, decorating personal igloo, and caring for pets called puffles. About every month, a seasonal or themed party was held for members to enjoy. This allowed for a unique experience where players could enjoy an altered world that was changed to match the party theme.

The business model of Club Penguin involved a free experience for all members, and a "Club Penguin Premium" experience that 10% of players opted to pay a monthly subscription for. In September 2015, Club Penguin began to show a decline when they closed down their German and Russian sites. In January 2017, Club Penguin announced their website would shut down on March 29, 2017. This was a tragic announcement. A final two-month "Waddle On Party" commenced during this time period for players to celebrate some final moments on their beloved game.

Club Penguin parties were a highlight of the user experience and game play. The goal of this analysis is to gain insights on the best year for partying in Club Penguin. Results of this analysis will allow former users to relive their glory days, and potentially identify trends correlated with Club Penguin's discontinuation in 2017.

Data

Data regarding Club Penguin is quite scarce. Fortunately, a Kaggle user named Saksham Nanda compiled a dataset with all Club Penguin Parties, their year, their start-date, their end-date, and some column with notes on various party features.

A few parties were missing start-dates and end-dates. Those rows were dropped during data cleaning. In addition, not all dates were formatted the same way, so some strings during data-cleaning were manually entered. New Years Parties were an issue when analyzing data year-by-year. For example, when a "New Years 2015 Party" would start December 2014 and wrap-around and end the first week of January 2015. To resolve this during data-cleaning, the "New Years 2015 Party" was split into two parties- one for the 2014 portion and one for the 2015 portion. Creating a duplicate party-row was a tradeoff for more convenient year-by-year data.

Notes on various party features, such as "Free-Items" provided an opportunity for a more rich text-analysis. However, a more simple approach was taken: generating a boolean variable "had_free_items" that indicated whether or not "Free-Items" said "None" for each party. Some parties had "special notes," such as "this was the first Club Penguin Party." This was simplified into a Boolean variable, "had_special_notes" to indicate if a party was unique in someway.

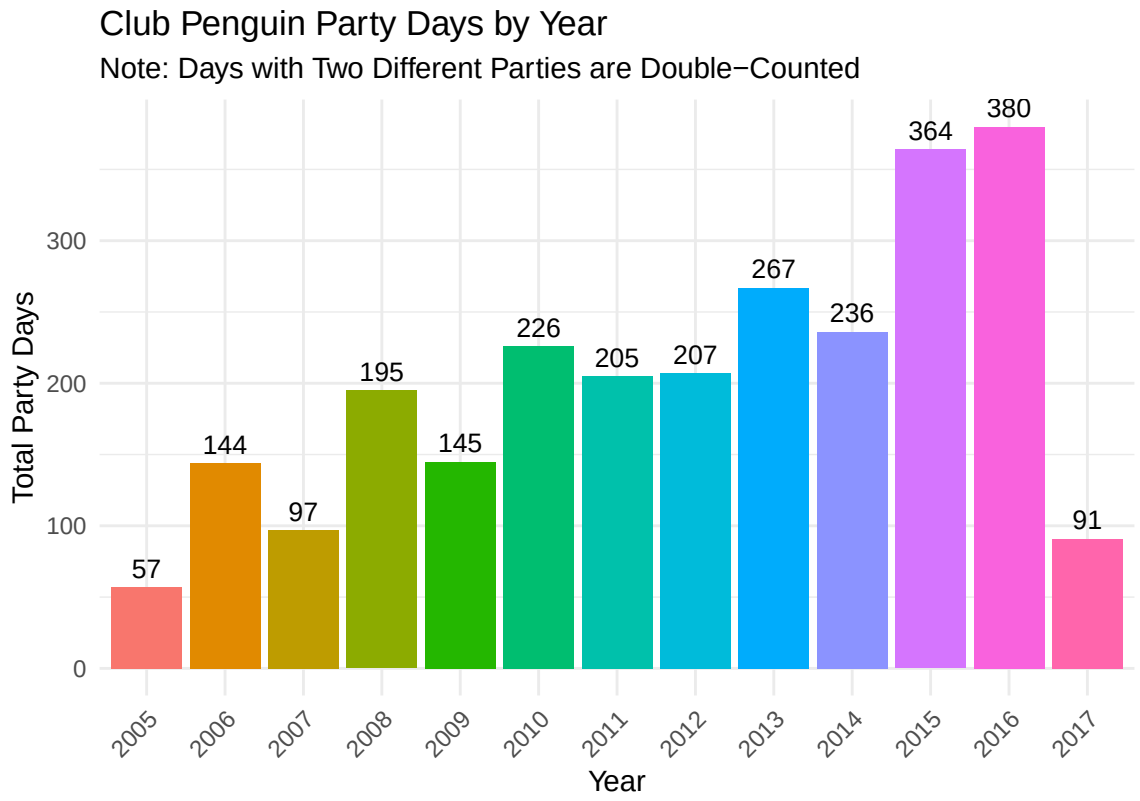
Unfortunately, no data was provided on the number of users in attendance at each party. This would have been an excellent metric to study, and a great indicator of how popular a particular party was. This analysis

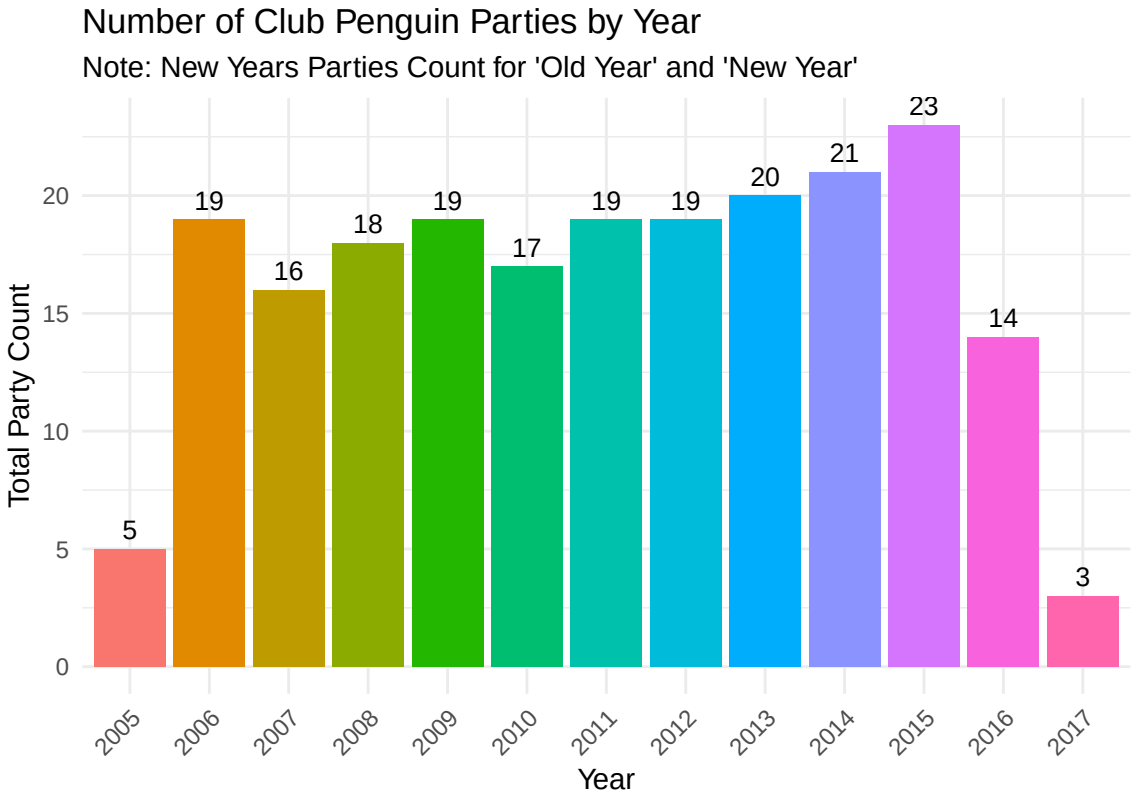
will focus mostly on party length in days, and number of parties per year. The dictionary below simplifies the variables to work with after data-cleaning:

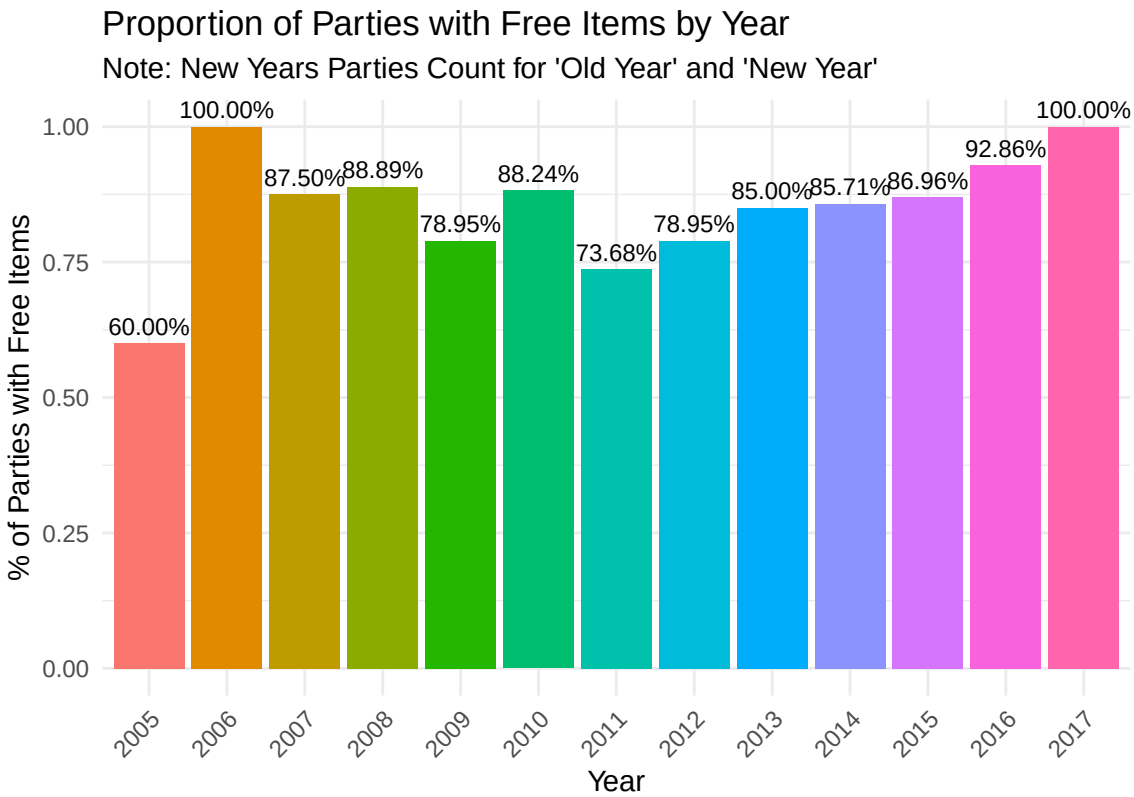
- $Party$: The party name.
- $Year$: Annual year when party occurred. Note that parties occurring in two years were split into duplicate party-names with different years.
- $StartDate$: The first date of the party, guaranteed to match $Year$.
- $EndDate$: The last date of the party, guaranteed to match $Year$.
- $TotalDays$: 1 + $EndDate$ - $StartDate$
- $HadFreeItems$: Boolean variable, indicates if party gave users free party-themed items.
- $HadSpecialNotes$: Boolean variable, indicates if original kaggle dataset recorded unique notes for this party.

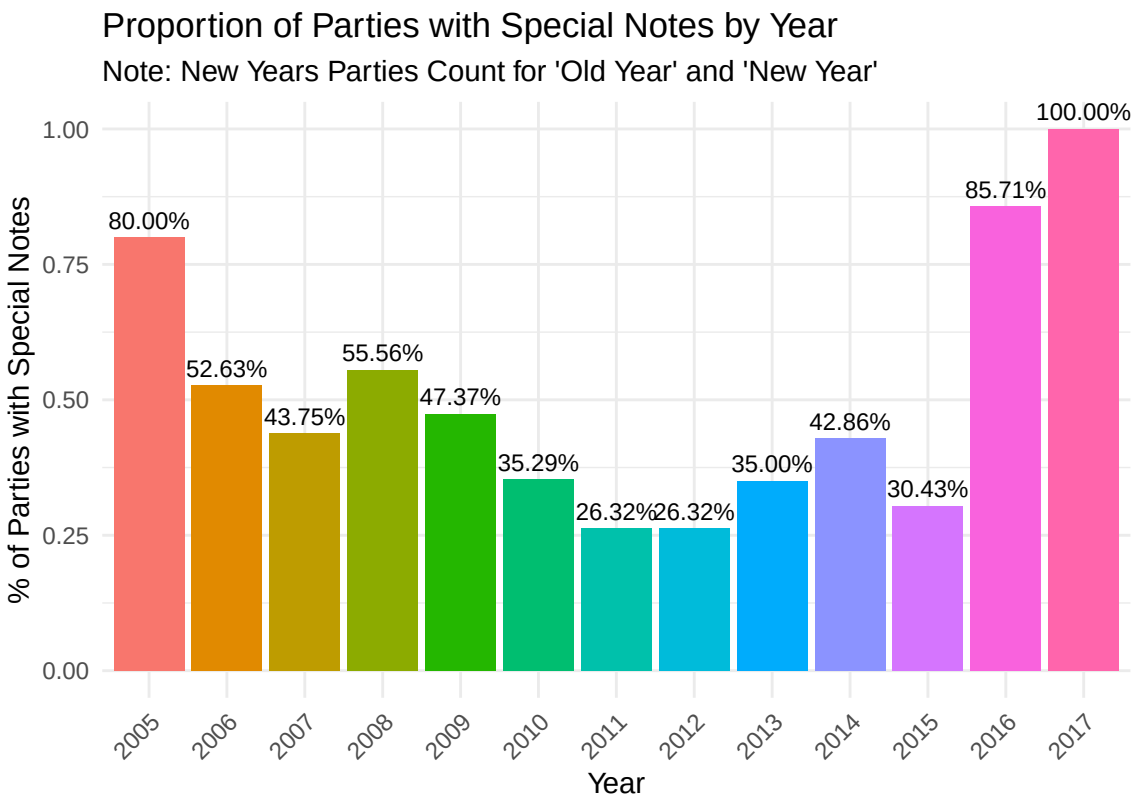
Exploratory Data Analysis

The following plots group "Party Rows" by year, and explore different distributions of key variables. Note that Club Penguin started in October 2005 and ended in March 2017. These years (2005 and 2017) are partial years, so some of their histograms counts are lower than other years.




Number of Club Penguin Parties by Year
Note: New Years Parties Count for 'Old Year' and 'New Year'
Total Party Count
Year
0
5
10
15
20
5
19
16
18
19
17
19
19
20
21
23
14
3
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017


Proportion of Parties with Free Items by Year
Note: New Years Parties Count for 'Old Year' and 'New Year'
% of Parties with Free Items
Year
0.00
0.25
0.50
0.75
1.00
60.00%
100.00%
87.50%
88.89%
78.95%
88.24%
73.68%
78.95%
85.00%
85.71%
86.96%
92.86%
100.00%
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017



"Club Penguin Party Days by Year" sums up all each party's length and adds them together. When ignorning the partial 2005 and 2017 years, the histogram shows an upward trend. Note that the year 2016 has more than 365 party days- this is due to overlapping Club Penguin parties on the same day. For example, the "Halloween Party 2016" occurred between 2016-09-28 and 2016-11-02 while the "11th Anniversary Party" occurred between 2016-10-19 and 2016-11-02. This has overlap during the end of Otctober.

"Number of Club Penguin Parties by Year" seems to average around 19 parties when ignoring years 2005 and 2017. However, note that 2016 only had 14 unique parties. This is notable because 2016 had the most "Party Days" out of all years, indicating game developers did not rotate parties as much as usual years.

No clear trend is shown by the "Proportion of Parties with Free Items by Year" histogram. The "Proportion of Parties with Special Notes by Year" histogram creates a "U shape." This shape is inline with the original data set, where the first and last parties of an annual holiday would have notes such as "First Annual Halloween party" or "Last Annual New Years party."

Methodology

The Methodology will drop years 2005 and 2017 from consideration when trying to analyze the best year(s) for parties in Club Penguin. The first and last year are partial years, have special parties surrounding them, and are simply not comparable to the full-length years.

The EDA above explores all columns in the limited dataset. However, the methodology will focus on two key variables: "Party Days Per Year" and "Number of Parties Per Year." Although some users may really like getting free items at their parties, the EDA shows no obvious outliers or trends for the full-length years. The "Special Notes" in the original data set doesn't necessarily indicate user satisfaction, just historical comments. Focusing methodology onto just two variables will allow for more reasonable interpretation than trying to combine all four of these variables into some alternate scoring system.

To evaluate candidates for potential best-party-years, we will perform two one-tailed z-tests for population

means at the $\alpha = 0.05$ level. These two tests will be repeated across all years, to identify potential best-party-years:

First One-Tailed Z-Test: H_0: The total "party days" for a given year are equal to the overall mean of total "party days" across all years. H_a: The total "party days" for a given year are greater than the overall mean of total "party days" across all years.

Second One-Tailed Z-Test: H_0: The total "number of parties" for a given year are equal to the overall mean of "number of parties" across all years. H_a: The total "number of parties" for a given year are greater than the overall mean of "number of parties" across all years.

Results

Table 1: Evidence of Above Average Annual Party Days

Year	Z-Score	P-Value
2006	-0.919	0.821
2007	-1.457	0.927
2008	-0.334	0.631
2009	-0.907	0.818
2010	0.021	0.492
2011	-0.220	0.587
2012	-0.197	0.578
2013	0.491	0.312
2014	0.135	0.446
2015	1.602	0.055
2016	1.785	0.037

Table 2: Evidence of Above Average Annual Party Ct

Year	Z-Score	P-Value
2006	0.150	0.440
2007	-1.090	0.862
2008	-0.263	0.604
2009	0.150	0.440
2010	-0.676	0.751
2011	0.150	0.440
2012	0.150	0.440
2013	0.564	0.287
2014	0.977	0.164
2015	1.803	0.036
2016	-1.916	0.972

Year 2016 in the "Evidence of Above Average Annual Party Days" table has a p-value of 0.037. The p-value is less than the alpha level, which suggests it is favorable to reject the null hypothesis. There is enough evidence to suggest that 2016 had more party days than the average year.

Year 2015 in the "Evidence of Above Average Annual Party Ct" table has a p-value of 0.036. The p-value is less than the alpha level, which suggests it is favorable to reject the null hypothesis. There is enough evidence to suggest that 2015 had more parties than the average year.

Based off the original methodology, which looks for any reason to favor a particular year as a "best-party-year," the results show 2015 and 2016 are good candidates for being the best-party-year in Club Penguin. The discussion section will look beyond the methodology and contextualize reasons to favor one year over the other.

Discussion

The original data set was missing a key metric for measuring the success of any party: attendance. Without attendance for each party, and eventually for each year, it becomes very difficult to measure what a "good year of partying" look like. This is why "number of parties" and "number of days with a party" were selected- they were the most clear annual numeric the dataset could provide. Even natural language processing on some of the "party notes" would be tough to do, as the notes were often empty or uninformative.

Given the lack of data, the methodology wanted to favor any reason to favor a particular year. One flaw of this approach can be found in the EDA, where year 2016 only has 14 parties when compared to the average of 19 for full-length years. Therefore, 2016 had only a few abnormally-long parties. The 1-tailed Z-test did not catch this underperformance. When picking between 2015 and 2016, the EDA (along with the methodology and results) shows that 2015 is the winner as the "best-party-year" based on the data provided.

Club Penguin had it's first company layoff in 2015 and shut-down in 2017. It's possible that developers wanted to keep their jobs in 2015 by adding more parties to the game (in hope of attracting more users). This is just speculation- it could also be a mere coincidence.

Of course, if someone really felt sentimental about the first party in 2005, or the final "Waddle On Party" in 2017, they would consider one of those as the best-party-year. Overall, this analysis attempts to quantify a very opinionated question on what makes a good year of partying.

References

- Wikipedia Page on Club Penguin: https://en.wikipedia.org/wiki/Club_Penguin
- Kaggle and Kaggle User Saksham Nanda: https://www.kaggle.com/datasets/mllion/club-penguin-2005-2017-all-parties
- ChatGPT: https://chatgpt.com/share/675fe9d9-6b30-8006-a643-73a3f57b95ab