NYP-NCHS Science & Technology Research Programme 2015/16

NCHS7 “Quickscope” Games Analytics

Done by: Yek Jin Teck Nicholas
Tan Yi Xiang
Shaun Tan
Lew Xuan Ming

Research Mentors: Mr Alvin Sebastian
Mr Sim Tze Jan

Teacher mentor: Ms Chney Chen Xiao Wei

Contents

I Abstract

Project 'Quickscope' is a project that utilizes data science in the form of game analytics to investigate player retention rates of a general player base by using samples of player data, to analyze various factors in to identify trends or traits among top performing players that the lower tier players do not share.

Using data sampling, we analysed the factors that influence Osu! Players, such as play time and total score and plotting the data into visual graphs to identify the traits they share in common.

Conclusions drawn from our analysis was that players who have surpassed a certain PP (Performance Points) threshold would most likely continue playing and remain active even over long periods of time as compared who did not achieve this threshold who tend to remain inactive for many months or quit the game entirely.

The significance of these results are that player retention rates can be significantly increased if the aforementioned conditions as well as crossing the PP threshold have been fulfilled. If players are assisted in reaching the threshold, Osu! would be able to steadily retain a portion of players and expand its player base further.

Figure 1. ***A screenshot of Osu! Gameplay***

II Introduction

Our Aim

Our project aims to improve the game industry via the application of data science. We aim to improve the game experience of players and hence, the retention rates for players. The tool we will use for our analysis is the IPython software to plot graphs and analyses data in order to identify trends among players.

The data parameters were obtained by the online rhythm game Osu!, in which the objective of the game is to clear game levels known as beatmaps, where a song would play and different rhythmic hit elements; hit circles, sliders and spinners would appear and players must correctly manipulate them in order to gain points. We would then gather data on 80 different players about their scores and time spent on the game, and analyze those factors to draw conclusions.

Problem Identified

The identified problem was that small companies without access to expensive advertising campaigns are unable to retain players due to steep learning curves or the lack of a striking start, and their games would slip into obscurity.

Osu! itself shares the problem of an extremely steep learning curve at the later stages that discourages players from continuing gameplay after a period of no visible progress. Given that it does not receive heavy media attention and that much of its popularity is based on word of mouth, the loss of players is much more keenly felt. Thus, by identifying the trends among top performing players, the reasons as to why they perform well and can be classified as retained players, Osu! can take action to supplement the weaker players and thus retain them.

Reason for the Choice of Issue

Game analytics is uniquely suited for students as it provides a gateway for data science, an industry of growing dominance with applications stretching from economics to mathematics. Experiencing these techniques allows the students to deepen their understanding of the vital science and be able to utilize it in the future.

Defining the Terms Utilized

Terms	What does it represent?
Retained Player/ Player Retention	A player that plays a game at least once a week is considered retained by the game, an active player in the game.
Performance Points (PP)	A system unique to Osu! to rank players according to their skill level, where achieving scores on difficult songs would award you PP in accordance to an unique algorithm.
Total Score	The cumulative sum of the scores of every song that players have attempted in Osu!, both ranked and unranked by Osu!.
Ranked Score	The cumulative sum of the score of every ranked song that players have cleared in Osu!, and directly influences their overall ranking.
DIFF	A factor determined by calculating Total Score – Ranked Score. The value represents how much more points worth a player has played failing and practicing a song compared to how many points the player has completed successfully. This value essentially represents how patient the player is.
Import	Imports functions from the written code
x=[ ], y=[ ]	A function that determines the X and Y values for the graph based on data inputted.
pl.plot(x,y)	A function that plots the x and y values of a graph in accordance to data inputted.
pl.Xlabel()	A function that allows written text to be included in the graph in order to provide labels for the data.
pl.show()	A function that calculates and displays a visual graph
Inactive	A player who has not logged into the game or has not completed a single level for more than a month.

III Investigation Process

Firstly, we worked to confirm the definition of game analytics and what it includes. We initially misunderstood what game analytics was and confused it with game analysis, and devoted resources pursuing the wrong endeavours.

Being a completely new field to us, we were then corrected by our mentors and then drastically shifted our research objectives.

We decided to find an existing game to collect data and do our research on.

Project Research

We had initially hoped to utilize data from UBISOFT (A popular game company) and construct a game and analyze its performance on the market. The data provided by UBISOFT would have provided us with a direction and an informed approach to the market, and we could craft a game that would not share the same problems in player retention. We created a game design document for our game [S.R.S.S], a 2-D platform action game featuring a robotic character using the Unity engine. It would have consisted of ten distinct stages and that would be selected and designed based on the data from UBISOFT which would indicate which particular environments and playstyles that generated the best response from players in the form of high activity. However, the plan was abandoned due to time constraints and no response from UBISOFT.

We then chose to utilize Osu! in our analysis, as they provide scores of player data for free on their webpage. Researching the retention rate of Osu! players, we formed a hypothesis:

Player retention would *more likely* occur if a player had <u>certain characteristics</u> and players would quit or play less due to the extensive difficulty curve in Osu!.

To support this theory, we first gathered data of 80 players each on two separate occasions, 13 September and 26 November using random sampling for factors such as their ranked score and play time to compare and draw conclusions from said data. Afterwards, we spent several weeks to devise a code that would analyze these factors and plot the figures into a graph before beginning our analysis and investigation. *Figure.2* below shows parts of our coding and research which we will explore more into later in the report.

```
In [44]: import pandas as pd
         df = pd.read_csv("OSUDATA.csv")
         import numpy as np
         import pylab as pl
```

```
In [1]: import pandas as pd
        df = pd.read_csv("OSUDATA2.csv")
        import numpy as np
        import pylab as pl
```

```
In [ ]: """PP to Inactive"""


        R = df.Rank

        NR = df.NationalRank
        ACC= df.Hitaccuracy
        TS = df.TotalScore
        RS = df.RankedScore
        TH = df.Totalhits
        PP = df.PP
        MC = df.MAXCombo
        PC = df.PlayCount
        LVL = df.Level
        IA = df.Inactive
        MP = df.MostPlays
        PT= df.PlayTime
        SS = df.SS
        S = df.S
        A = df.A


        x = [PP]
        y = [IA]
        pl.plot(x, y, 'ro')
        pl.xlabel('Performance Points')
        pl.ylabel('Inactive (Months)')
        pl.show()

        #0-4kpp = Most unretention | 4-5kpp = obvious less retention | >5k pp players mostly active. So what's the diff between
```

```
In [1]: import pandas as pd
        df = pd.read_csv("OSUDATA3.csv")
        import numpy as np
        import pylab as pl

        TS = df.TotalScore
        RS = df.RankedScore
        IA = df.Inactive
        PP = df.PP

        x = [PP]
        y = [IA]
        pl.plot(x, y, 'ro')
```

Figure 2. ***A picture showing parts of our code***

Project Timeline

11/3/2015 **(First Mentor Meeting)**	We met up with our mentors for the first time and we discussed what game analytics was, and how it was different from game analysis, given that the two are commonly confused with each other.
3/6/2015 **(First Group Meeting)**	We discussed the ten factors that we would base our research around, and forge our game utilising the data from UBISOFT factored into it to uncover the reasons for low player retention rates and then confirm our hypothesis by launching the game into the market and checking its market progress.
18/7/2015 **(Second Mentor Meeting)**	We realized that we had completely misunderstood what game analytics meant and had confused it with game analysis, and then greatly altered our plans, and shifted our goals to gathering player data from Osu! and analyzing them to uncover traits that would serve the purpose of increasing retentions rates of players.
23/11/2015 **(Third Mentor Meeting)**	On this occasion we discussed with the mentors how to utilize the data we have gathered since the previous session such as to form graphs or using various data analysis tools such as Google's Big Query, or various others.
11/7/2015 **(Second Group Meeting)**	We went over to Nicholas' house to share and learn things that we have thought about such as the hypotheses that we have come up with as well as the basic programming skills mainly learning the programming language 'Python' that were needed to operate the system for our project, IPython.

	While completing the code that we would use to analyse and plot our gathered data.
18/12/2015 (Third Group Meeting)	Here we utilised the code that we created the previous week to plot various factors against each other to analyse the various graphs created, and after comparing differing factors we successfully identified distinctive traits among top players.
22/12/2015 (Fourth Mentor Meeting)	We shared our results with the mentors and discussed with them how to correctly interpret and understand the graphs, while receiving confirmation that our data was accurate and concise.

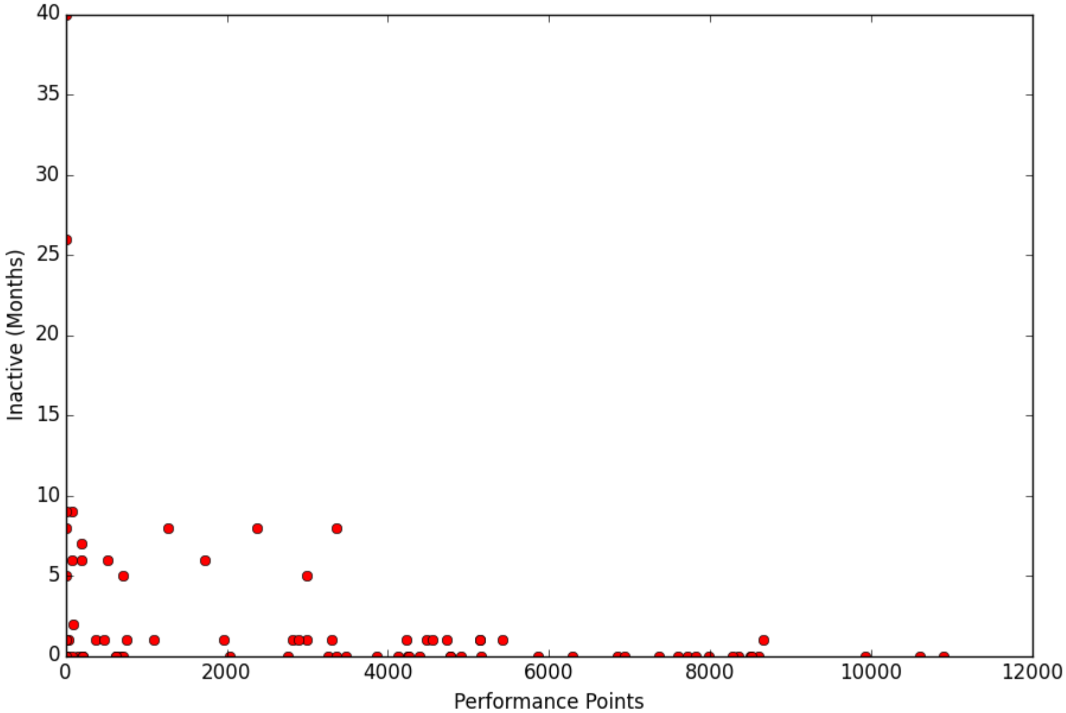
IV Findings and Results

Based on the data that we have gathered, as well as trial and error, we have noticed a trend where players that have achieved 5000 PP or more will almost certainly be retained and raise their log in rates to near daily and remain active in the game for extended periods of time.

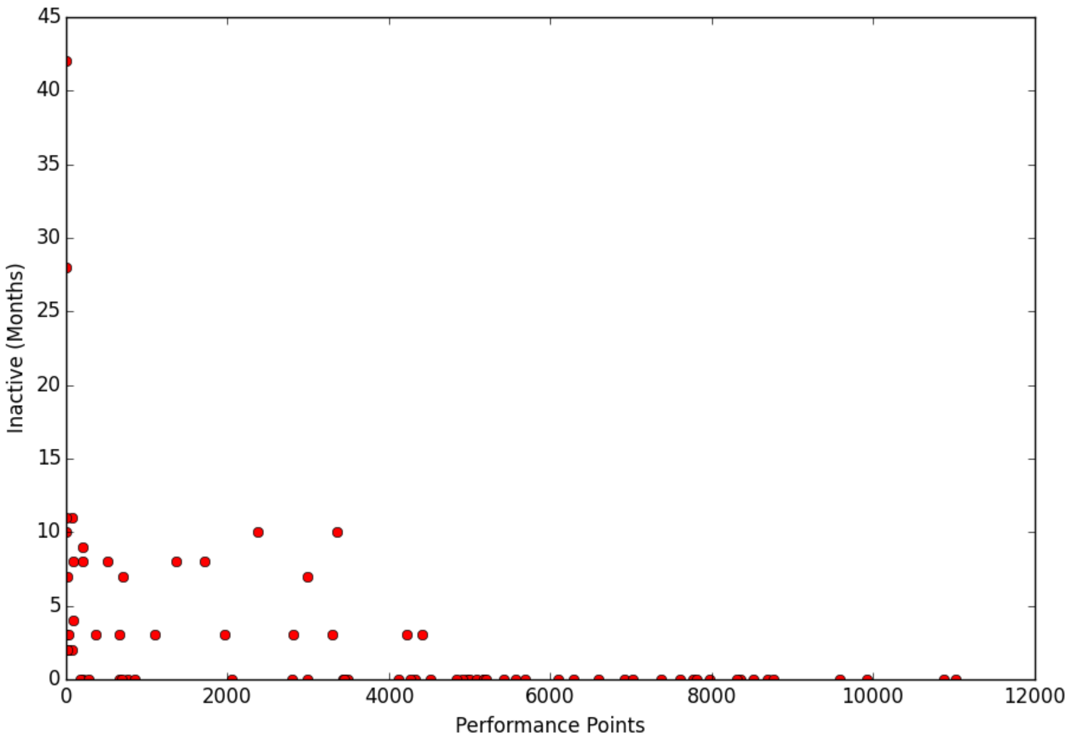
Among the players that have achieved this threshold, the data indicated that the DIFF among players scaled exponentially against PP, meaning that the players with more patience and consistency in playing were more likely to reach said PP threshold and remain more active as they can constantly improve their skills from that point onwards.

We plotted the graphs as such:

```
x=[Variable1]
y=[Variable2]
pl.plot(x,y, ‘ro’)
pl.xlabel(‘Variable1Name’)
pl.ylabel(‘Variable2Name’)
pl.show()
```



<u>Figure 3.0 *1st Data Collection of 80 players*</u>



<u>Figure 3.1 *2nd Data Collection of same 80 players*</u>

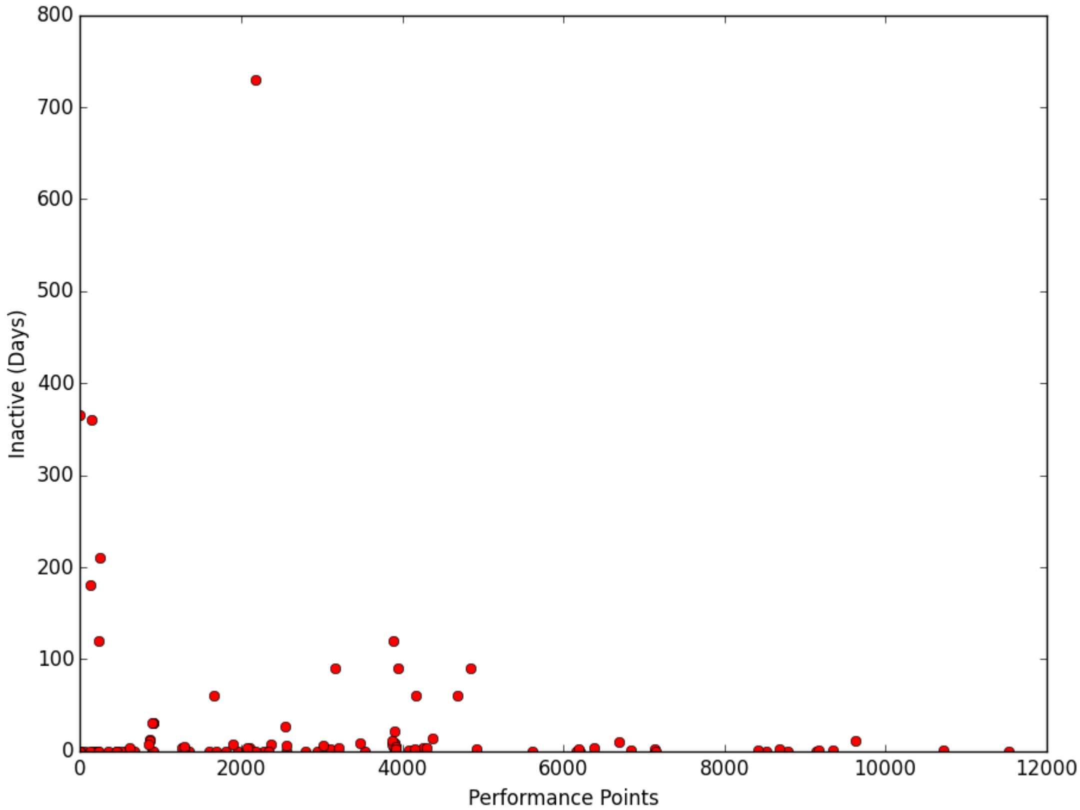


Figure 3.2 ***PP against Inactivity (Days) New sample of 103 Players***

In Figure 3.1 we plotted PP against Inactivity with our first set of data with 80 players. It is observed that inactive players dramatically drops at ≈5000PP. Players below 5000PP are more likely to be unretained (inactive), and players >5000PP are shown to be active and have very low (close to none) numbers of inactive players.

This trend is further supported by Figure 3.2 and 3.3.

Figure 3.2 shows the same 80 players at a later date. The trend is once again observed. Players >5000PP are still shown to be active and retained, whereas players <5000PP are more likely to be inactive.

Figure 3.3 show the plot with a new sample of 103 Players. It once again shows the same trend. Players >5000PP are more active compared to Players <5000PP. All of the players >30 days inactive are shown to be <5000PP Players.

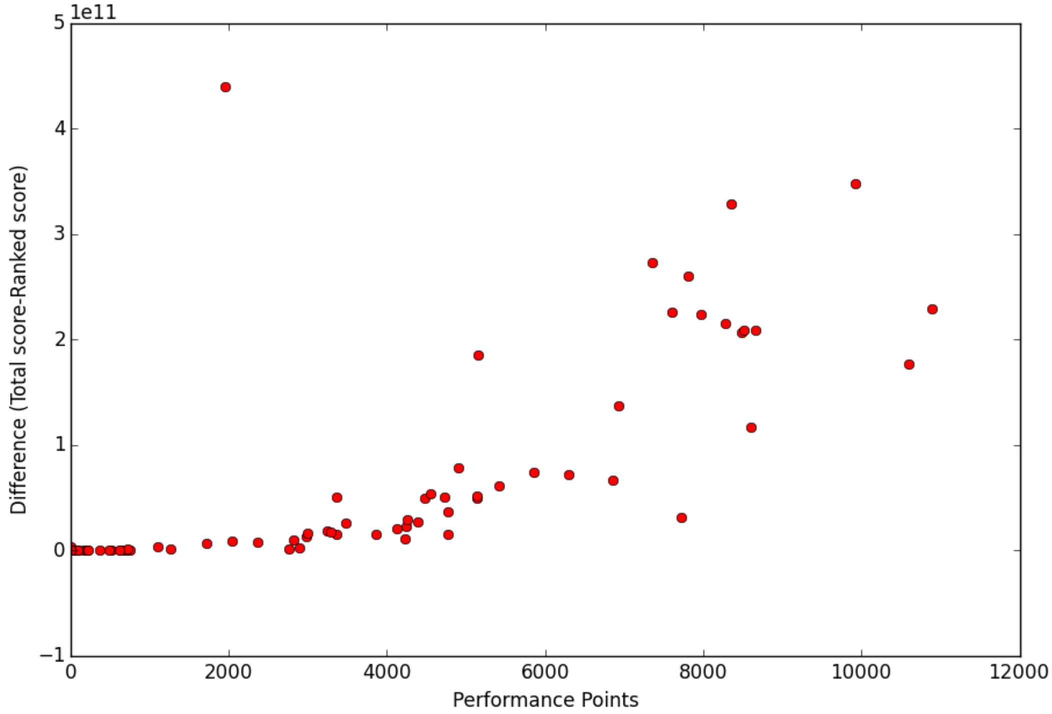


Figure 4.0 *PP against DIFF* [1st Data Sample]

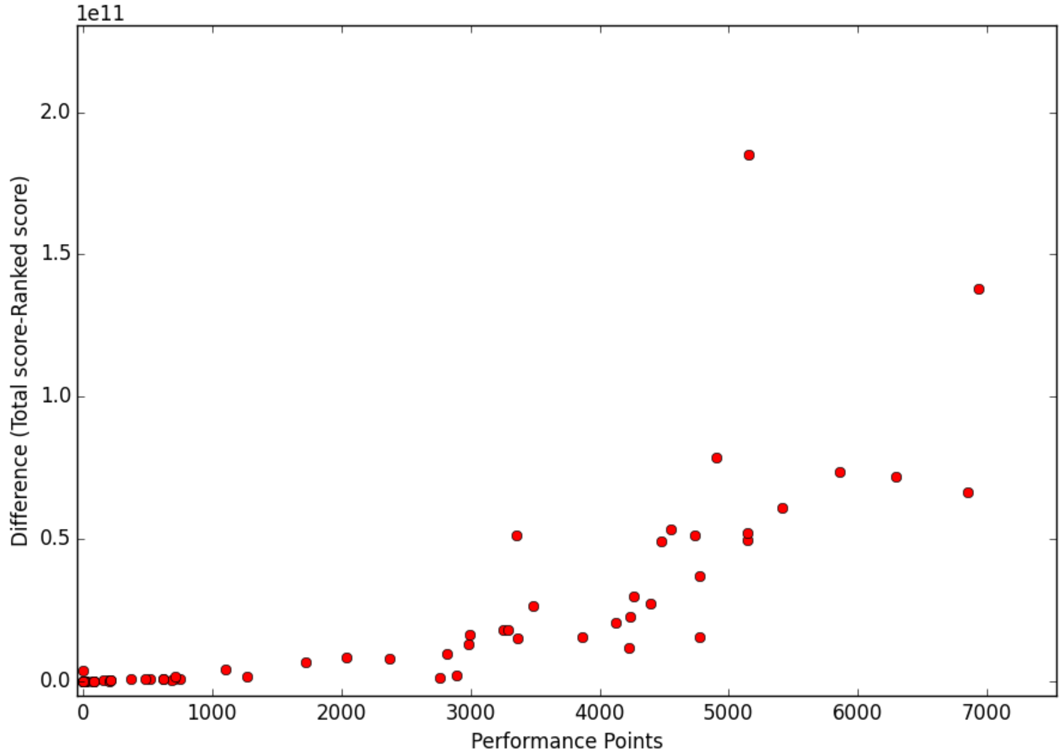


Figure 4.1, Same graph of 4.0 but close up.

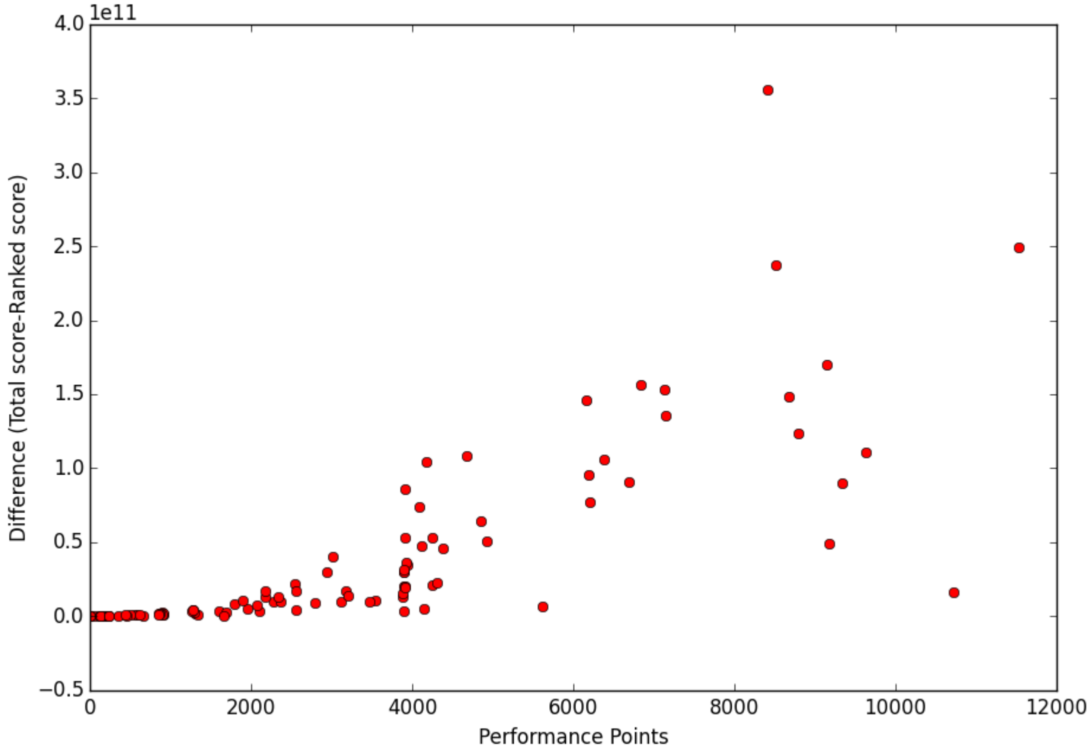


Figure 4.2 PP against DIFF Close Up [Data collection 2]

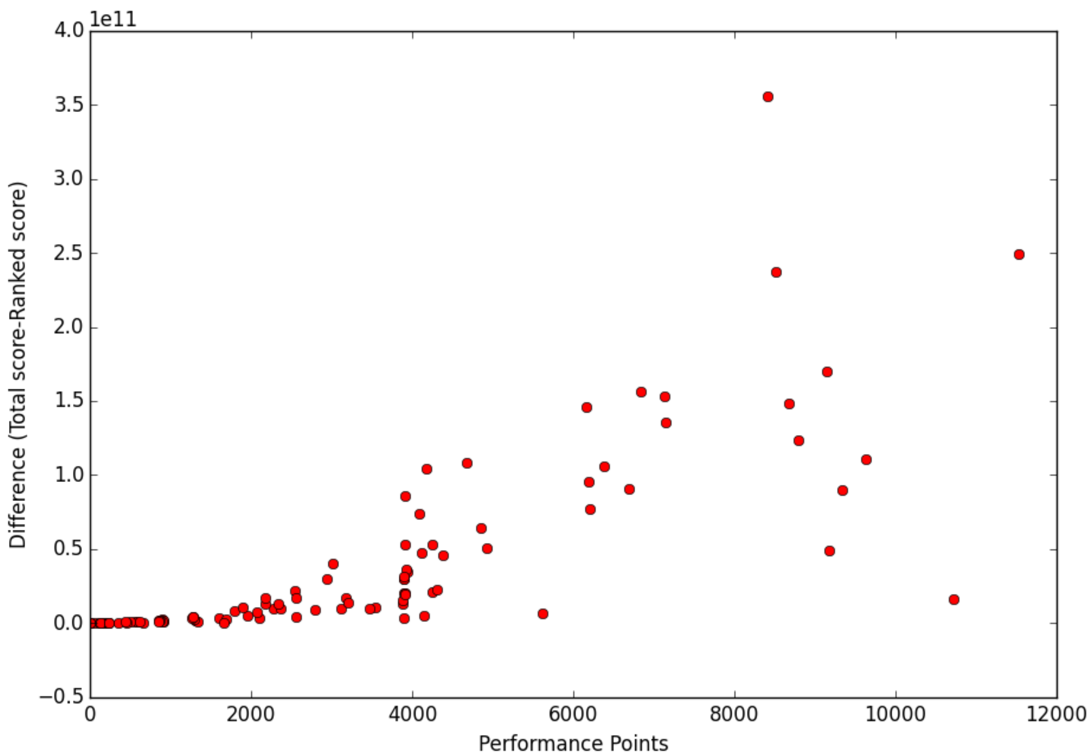


Figure 4.3 PP against DIFF [Data Collection 3]

In Figure 4.0, We can see a trend whereby the more active players (≥5000PP) are seen to have this particular trait: Most of these ≥5000PP players have a DIFF value of 0.5×10^{11}.

In Figure 4.1, the cut-off point (0.5×10^{11} DIFF) is much more clearly seen close up.

Figure 4.2 shows the same trend again. Most players ≥5000PP are shown to have 0.5×10^{11} DIFF.

Figure 4.3 shows the trend again with a larger sample size of 103 players. Once again, most players score 0.5×10^{11} DIFF if their PP is ≥5000PP.

It is noticed that there are some “Special Cases” whereby ≥5000PP players may have a low DIFF value. This means that the player is extremely skilled whereby his Total Score is very close in value with Ranked Score, which means that for these Special Cases, the player is able to play well consistently without needing as much practice or retries that those of similar PP.

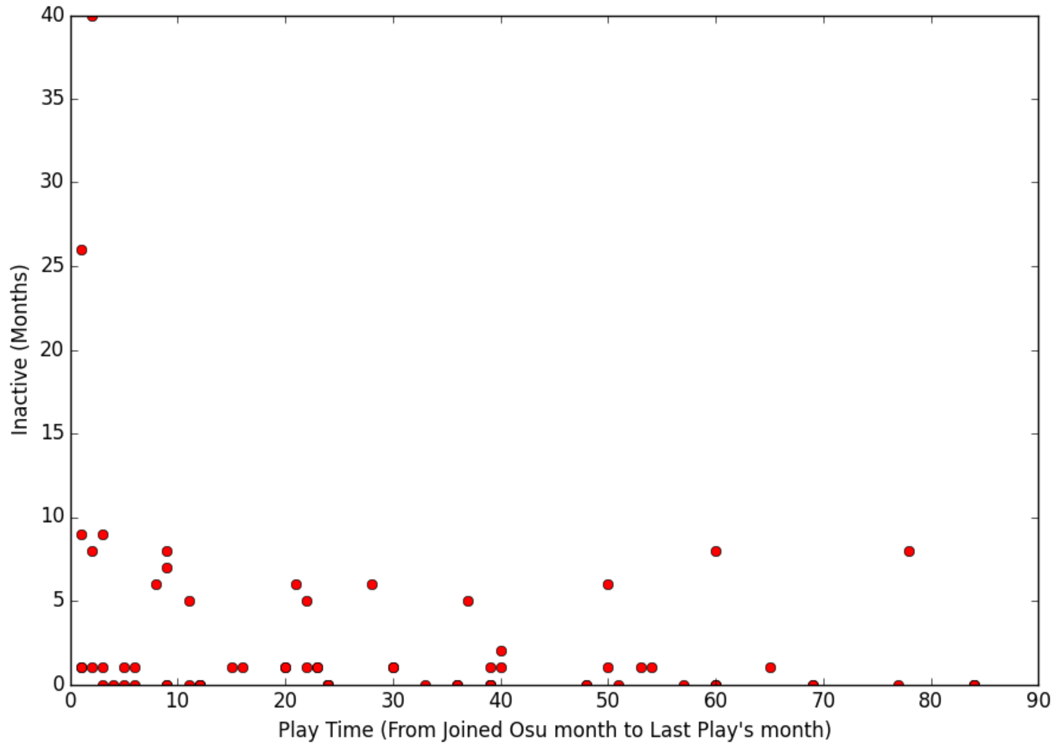


Figure 5 Playtime against Inactive [1st collection]

In Figure 5, we can conclude that the general trend is that no matter how long a player plays Osu! (for from their join date to last login), Players can still be inactive.

Essentially, a long-membership player has the same likelihood as players who have just joined or are fairly new to be inactive.

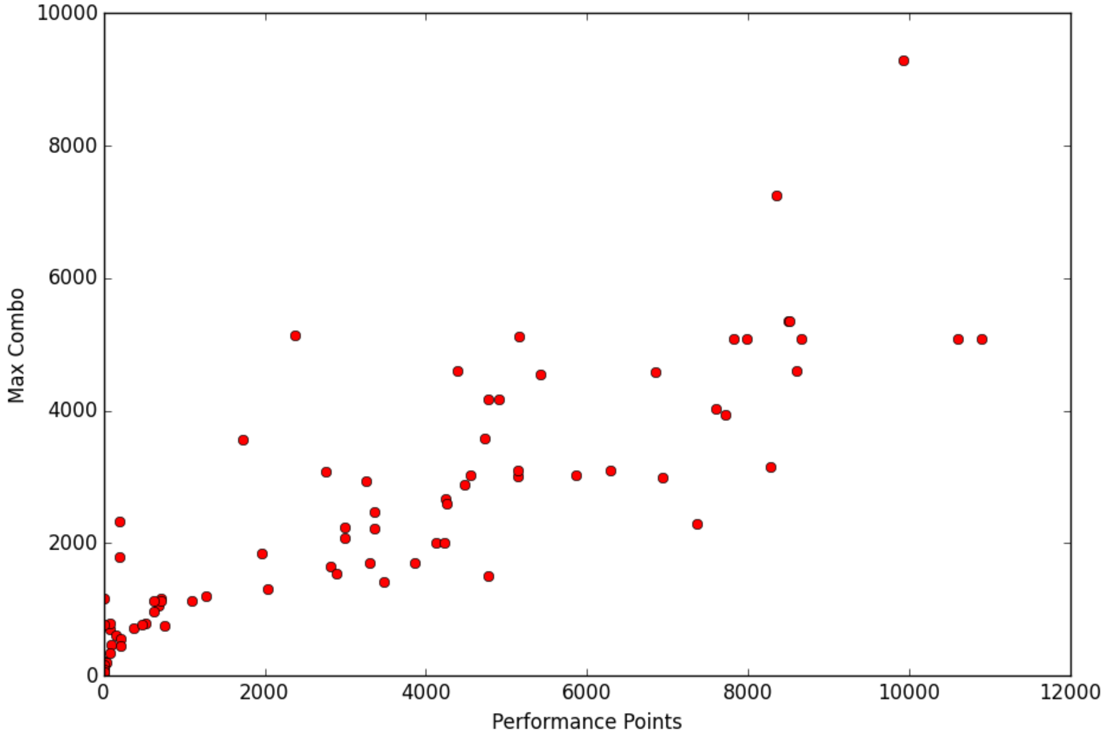


Figure 6 PP against Max Combo

In Figure 6, we can see that the general trend is that players ≥5000PP have the common trait of having ≈ 3000 or more in their Maximum Combo value apart from 1 or 2 exceptions.

It is noted that to have high Max Combo values, a player must be patient and very skilled to endure and play long songs to reach such values. To put it into perspective, the cut-off point 3000 Max Combo is equivalent to 3000 on-beat taps on the keyboard.

V Analysis of Findings

We utilized IPython and the Pandas library to plot the data into graphs and analysed them in that form as it was recommended by the mentors over software such as Google's Big Query and it would provide us students with a simple and easy to use method for data analysis. Although it was a lot of trial and error, we managed to come up with some possible conclusions.

Our 1st and 2nd data collection of 80 players and 3rd data collection of 103 players from the lower, middle and upper ends of the player rankings through random data sampling ensures a level of accuracy in the form of reduced human bias in the data we gather.

From Figures 3.0 to 3.2, we can see that almost all players that are ≥5000PP the players whom are the active members of Osu! and are retained (active). 5000PP is the benchmark whereby players who pass it are most likely to be very active players.

From Figures 4.0 to 4.3, we can see that to achieve ≥5000PP, players need to be patient and consistent. The figures show the majority of players with 0.5×10^{11} DIFF having ≥5000PP. A higher DIFF value means that a player has re-tried many songs over and over again (which causes a high Total Score and a significantly lower Ranked Score) *hence*, a player with high DIFF value is more patient.

From Figure 5, we also find that players regardless of how long their membership is, can still be unretained (inactive). Long membership does not mean a player will be retained. It is rather, the skill and patience a player has the makes the difference.

From Figure 6, we found that ≥5000PP players (apart from 1 or 2 exceptions) had at least 3000 MAX COMBO. This further shows that almost all ≥5000PP players are *patient and consistent* in gameplay. This is due to players having to persevere and needing skill to be able to reach a combo of 3000.

VI Conclusion

In conclusion, we believe that players that have crossed the 5000 PP threshold will most likely be retained and remain active in the game. In order for Osu! to increase player retention numbers, Osu! should make efforts towards *assisting players* to reach the said threshold. The players with more patience were more likely to reach the score threshold and remain more active.

We would suggest Osu! implement a checkpoint system to help out the less patient and skilled players on hard songs with certain chokepoints where the player would always fail (then giving up prematurely) due to lack of skill.

VII Acknowledgement

We would like to thank our teacher mentor, Miss Chney Chen and our research mentors: Mr Alvin Sebastian and Mr Sim Tze Jan as they have assisted greatly in our project in the form of constructive criticism and guidance when we were unsure about the project, especially during the initial stages when we were trying to tackle the problem of a platform for game analytics to shine. We would also like to thank Mr Tan and Mdm Chow, our principal and vice-principal for supporting this program and allowing us this opportunity to develop ourselves as students.

VIII Bibliography

http://osu.ppy.sh.com
http://www.thedailycrate.com/wp-content/uploads/2014/04/screenshot007-2.jpg
https://www.continuum.io (Ipython/Jupyter notebook function)