King’s Super Math Saga experiment

The purpose of this report is to provide the Head of Studio the conclusions about the Super Math Saga experiment performed in 2017 between May, 4th and May 22nd, and highlight the difference in the behaviour of the two groups object of the A/B test. Data for the period April 20th - May 3rd were also available, and the analysis includes all the data available.

# Summary of portfolio

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  | **20th Apr – 3rd May** | | **4th May – 22nd May** | |
| Table 1 | **Total** | **Group A** | **Group B** | **Group A** | **Group B** | **Group A** | **Group B** |
| Users[[1]](#footnote-1) | 204,103,781 | 163,317,198 | 40,786,583 | 63,897,583 | 15,952,936 | 99,419,615 | 24,833,657 |
| Unique users[[2]](#footnote-2) | 10,331,056 | 8,265,610 | 2,065,445 | 6,776,037 | 1,692,085 | 8,265,610 | 2,065,446 |
| Gamerounds | 2,684,253,732 | 2,152,940,504 | 531,313,228 | 842,316,960 | 210,156,002 | 1,310,623,544 | 321,157,226 |
| Purchases | 6,283,916 | 4,982,140 | 1,301,776 | 1,955,183 | 490,943 | 3,026,957 | 810,833 |
| Games per user[[3]](#footnote-3) | 259.82 | 260.47 | 257.24 | 124.31 | 124.20 | 158.56 | 155.49 |
| Conversion[[4]](#footnote-4) | 0.23% | 0.23% | 0.25% | 0.23% | 0.23% | 0.23% | 0.25% |
| Purchase by client[[5]](#footnote-5) | 0.6083 | 0.6028 | 0.6303 | 0.2885 | 0.2901 | 0.3662 | 0.3926 |

Table 1 shows summary data for the portfolio since inception date for the most important measures considered in this report.

The assignment date for some users is before the start of the experiment and the pre-experiment period data are useful to highlight differences in Group B with the period of the experiment. This report refers to the period 20th Apr - 22nd May as the *entire period*, namely the whole period for which we have data. The name *experiment period* refers to the period 4th May - 22nd May.   
Group B is the experimental group made of 20% of the total sample. The overall number of games per user is 259.52 with a conversion of 0.23%. The difference in gamerounds per user between group B and group A is -3.07 in the period of the experiment, and the difference in conversion is 0.02%. The number of days in row users play the game is another valid measure for the engagement. As sanity checks, it is important to notice the number of users (unique and total) reflects the assumptions of the experiment (80%-20%), and there is no user id in common between the two groups[[6]](#footnote-6). Before the experiment on average each user plays 124 gamerounds, during the experiment this number jumps to more than 155 but the average number for the control group (group A) is higher than the experimental group (group B), this hints that the experiment did not result in increasing the number of games played by user for group B, the conversion instead improved of 0.02% and the purchase by client of 0.0264.

The [*EDA (Exploratory Data Analysis)*](#_EDA_–_Exploratory) section, will present the data exploration, then the [*Analysis*](#_Analysis) *section* highlights the methodology used to formulate the conclusions, and finally, the [*Insight*](#_Insights_1) section summarises the key indications for King emerging from this study.

## EDA – Exploratory Data Analysis

|  |  |
| --- | --- |
| Figure 1 | Figure 2 |
| A screenshot of a video game  Description generated with high confidence | A screenshot of a cell phone  Description generated with very high confidence |
| Figure 3 | Figure 4 |
| A picture containing screenshot  Description generated with very high confidence | A screenshot of a cell phone  Description generated with high confidence |

This section is to give the reader a deeper understanding of the data. There are two groups of users that the report calls A or control and B or experimental. The users in group B are subject to a change, that is the object of the experiment, and therefore their engagement should be higher. The KPI that are proposed for the engagement are those defined in the bottom rows of Table 1 namely: games per user, conversion and purchases per user. Table 1 shows that the value for the games per user does not improve for group B compared to group A (the figure is in red for this purpose). Figure 1 and Figure 2 show a boxplot for the period before the experiment and after. The difference in the median value is minimal before the experiment but after is decimal of percentage points. Figure 3 and Figure 4 show a density plot for the same variable. We can observe that while before the experiment the purchase per user was between 0.029 and 0.031 with a probability not below 0.9 during the experiment such probability becomes much lower and the interval widens, therefore the purchase per user will be between 0.030 and 0.036 with a probability of at least 0.9.

Afterwards, it is important to analyse the number of gamerounds and purchases by date. A dual-axis chart is useful in this situation since there is a big difference in the absolute numbers between the two groups (Figure 5).

|  |  |
| --- | --- |
| Figure 5 | Figure 6 |
|  |  |

Figure 5 shows the absolute number gamerounds by date between the two groups. Figure 6 shows the absolute number of purchases by the group. Alongside with the absolute number, it is important to look at relative figures defined with four ratios using the engagement metrics at disposal (gamerounds and purchases). The average engagement is the denominator of this ratio. Since the number changes if the period considered[[7]](#footnote-7) changes, this report considers four ratios defined as below[[8]](#footnote-8):

|  |  |  |
| --- | --- | --- |
|  |  |  |

Figure XX shows the gamerounds ratio entire, Figure XX shows gamerounds ratio experimental, Figure XX shows purchases ratio entire and Figure XX shows purchases ratio experimental

The key findings are:

1. There is not much change in the shape of the curve for the gamerounds, and for purchases, the benefit of the treatment does not seem very noticeable
2. The gamerounds in the period of the experiment for group B is not better compared to the group A
3. The conversion during the period of experiment for group B looks better than group A but the difference is 0.02% at its maximum

It is also possible to take into consideration the day of the week and check if there are one or more days that present better value for the engagement. For this purpose, the report considers the day of the week for each day in the period and calculates the average for gamerounds and purchases. Then the total value for gamerounds and purchases for each calendar day is divided by the average of the correspondent day of the week. This ratio would give information if one day of the week showed a higher number of users. This document repeats the analysis considering the average in the period before the experiment[[9]](#footnote-9) and during the experiment[[10]](#footnote-10). The seasonality index for purchases only is showed in Figure 7. Looking at these charts, it is evident that there is not much difference between group A and group B. Looking at the seasonality index calculated by day of the week before the experiment and after, displayed in Table 2, it is evident that there is no bigger variation.

|  |  |
| --- | --- |
| Figure 7 | Table 2 |
|  |  |

Until now, the report considered the engagement basically as a share of gamerounds and conversion, but another important measure is the streak, defined as the number of consecutive days in which users play the game. This report considers a 5-days streak as well. Table 3 summarises the results. Again it is evident that there is not much difference between the control and experiment group. It is not significant to check more streak because the analysis would be meaningless since the 5-days did not bear results. Considering a streak with less day is meaningless too.

Table 3

## 

## Analysis

This section explains the methodology that is used to examine the data and draw conclusions. The language used is technical, but everything is explained to make this document readable even by a non-technical person.

As stated before in the paper, the control group (group A) has 80% of the users while the experimental group only the remaining 20%, therefore using ratios will allow us to compare the two groups in an easy way. This paper stated also that data pre-test are present. These data are important because it is possible to see a difference between prior the experiment and after the experiment and this is meaningful for group B but also for group A because it will tell if there is another factor that is acting that could not be singled out just looking at group B.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Pre-test (PT) | Test (T) |  |
| Group A – control | KPIA-PT | KPIA-T | KPIA-T - KPIA-PT : no difference |
| Group B – experiment | KPIB-PT | KPIB-T | KPIB-T - KPIA-PT > 0 & significant |
|  | KPIA-PT – KPIB-PT : no difference | KPIB-T - KPIA-T > 0 & significant |  |

It is important to establish if these differences are real or due to chance. In the case of the experiment, the outcome of the events is binary (purchase/non-purchase, game/no-game); therefore, the report uses the *binomial distribution*. This section considers only purchases, afterwards in the following section, the conclusions will be drawn. It is possible to define four metrics: Xexp sales in the experimental group (B), Xc sales in the control group (A), Nexp unique users in the experimental group and Nc unique users in control group. Therefore, the probability that users in the experimental group buy some level is while in the control group is . It is crucial to establish if the difference observed ( ) is real[[11]](#footnote-11). So it is possible to build a confidence interval for the true difference between the probability of converting in the experimental sample and the control sample. This paper chooses a 95% confidence interval. At this point, it is important to define two hypotheses, H0 the null hypothesis that states that the true difference is 0[[12]](#footnote-12) and the H1 alternative hypothesis that states that the difference is different from 0. In building the confidence interval, it is essential to calculate the probability and standard error of the pool[[13]](#footnote-13).

## Insights

The difference between group A and group B is very little and although it is statistical significant it is very small and it seems the experiment did not help to differentiate significantly group B compared to group A. It is important the practical significance, it means that the change has a real impact for the business and therefore it is worthy to be implemented. As seen in this report the engagement is substantially the same for the two group of users. The report also calculated the streak (the details of the calculations are in the attached notebook) but again no difference is found between the two groups. On this basis this paper advocates not to spending time and effort implementing the change that the experiment wanted to test.

1. Number of user\_id present in the activity table [↑](#footnote-ref-1)
2. The breakdown does not add up because of double counting [↑](#footnote-ref-2)
3. The denominator used is the unique number of users [↑](#footnote-ref-3)
4. Average values calculated as *total purchases* divided by *total number of games* [↑](#footnote-ref-4)
5. Average values calculated as *total purchases* divided by *number of unique users* [↑](#footnote-ref-5)
6. This has been checked with query XXXX – see attachment of query details [↑](#footnote-ref-6)
7. Entire period from which data are available (20/04 – 22/05) or experimental period (04/05 – 22/05) [↑](#footnote-ref-7)
8. I.e. Gamerounds entire period = (Total gameroundsi/(Avg gamerounds entire period) with *i* being each day in the entire period (20/04 – 22/05) and the average, the denominator, is calculated over this period [↑](#footnote-ref-8)
9. 20/04 – 22/05 [↑](#footnote-ref-9)
10. 04/05 – 22/05 [↑](#footnote-ref-10)
11. The word used in statistic is *statistically significant* [↑](#footnote-ref-11)
12. It is normally distributed with mean 0 and standard deviation equal to the standard deviation of the pool [↑](#footnote-ref-12)
13. [↑](#footnote-ref-13)