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 | **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 test. This report refers to the period 20th Apr - 3rd May as the *pre-experiment period*, namely the period before the experiment began. 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 essential 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 research on average, each user plays 124 gamerounds; during the experiment, this number jumps to more than 155. However, the average amount 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 the 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 critical indications for King emerging from this study.

## EDA – Exploratory Data Analysis

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

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 KPIs that 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 test, but after it 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 research such probability becomes much lower, and the interval widens; therefore the purchase per user will be between 0.030 and 0.036 with a likelihood of at least 0.9.

Furthermore, it is essential 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 | Figure |
|  |  |

Figure 5 shows the absolute number gamerounds by date between the two groups. Figure 6 shows the number of purchases by the group. Alongside with the value, it is relevant to look at relative values defined with four ratios using the engagement metrics at disposal (gamerounds and purchases). For this case the conversion (the number of sales divided by the number of gamerounds) and purchases per user (number of purchases by distinct users) are useful.

|  |  |
| --- | --- |
| Figure | Table |

The proportion of the sales per user per group is calculated by each day for each group, and Figure 7 displays it. Table 2 gives the average of purchases by a user for the period before and during the test. For the chart in Figure 7, the average of purchases per user is calculated by each day for each group.

The key findings are:

1. The gamerounds in the period of the experiment for group B are not better compared to group A
2. The conversion during the period of the research for group B looks better than group A, but the difference is minimal (0.02% on average)
3. Purchases per-users improve from 3.08% to 3.27% on average for group B

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 considers an analysis of the average in the period before the experiment[[7]](#footnote-7) and during the experiment[[8]](#footnote-8). The seasonality index calculated using the averages by day of the week presented in Figure 7. Looking at the chart, it is evident that at the commencing of the experiment, there is a 0.2% improvement for group B, but then it quenches towards the end. Looking at the seasonality index calculated by day of the week before the experiment and after, displayed in Table 3, it is evident that there is no bigger variation.

|  |  |
| --- | --- |
| Figure | Table |
|  |  |

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 4 summarises the results. 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 not very impactful on engagement.

Table

## 

## 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 also stated that data pre-test is 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 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[[9]](#footnote-9). Therefore 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 bases its conclusions on 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[[10]](#footnote-10) 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[[11]](#footnote-11).

## Insights

Table

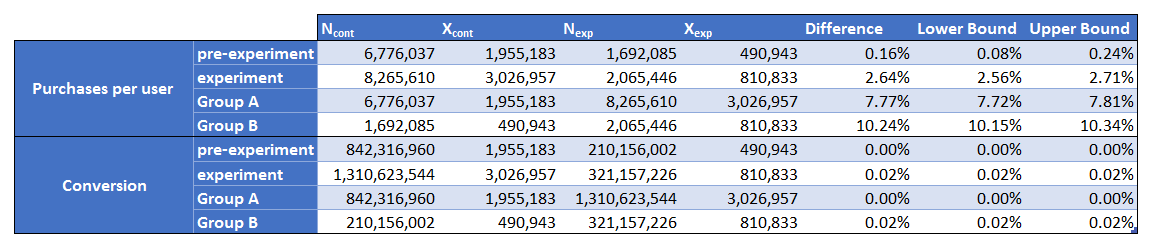


Table 5 shows the hypothesis test on the purchases per user, considering four cases: before the experiment, after the experiment (comparison between group A and group B) and group A, group B (comparison between the period before and after the experiment). Before the experiment, the purchases per-users was equal between group A and group B, but after the research, there is a statistically significant difference of 2.64%. Looking at group A, whose users did not receive the treatment, between the period before and after the test, there is an improvement of 7.77% that is statistically significant. For the users in group B, the increase was of 10%; therefore, they performed 2.64% better than group A. Such a change in purchases per user justifies implementing the changes tested. The report calculated conversion and proposed analysis on the streak (the details of the calculations are in the attached notebook) but, in these cases, there is no difference between the two groups (for conversion look at Table 5.)

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 the *total number of games* [↑](#footnote-ref-4)
5. Average values calculated as *total purchases* divided by the *number of unique users* [↑](#footnote-ref-5)
6. This has been checked with query 5 – see attachment of query details [↑](#footnote-ref-6)
7. 20/04 – 22/05 [↑](#footnote-ref-7)
8. 04/05 – 22/05 [↑](#footnote-ref-8)
9. The word used in the statistic field is *statistically significant* [↑](#footnote-ref-9)
10. It is normally distributed with mean 0 and standard deviation equal to the standard deviation of the pool [↑](#footnote-ref-10)
11. [↑](#footnote-ref-11)