

Kelompok 3 DA1

Nama Anggota	ID Participant
Audrey Simae Anaa Dachi	KM_G2DA1019
Ulfah Nur Oktaviana	KM_G2DA1379
Stevani Dwi Utomo	KM_G2DA1289
Laras Sekar Kinasih	KM_G2DA1199
Ruach Sakadewa	KM_G2DA1109

Cookie Cats

Cookie Cats is a hugely popular mobile puzzle game developed by Tactile Entertainment. It's a classic "connect three" style puzzle game where the player must connect tiles of the same color in order to clear the board and win the level. It also features singing cats.

As players progress through the game they will encounter gates that force them to wait some time before they can progress or make an in-app purchase. In this project, we will analyze the result of an experiment where the first gate in Cookie Cats was moved from level 30 to level 40. In particular, we will analyze the impact on player retention and game rounds

Part 1

Goal/Objective

Increase Retention Rate

General Problem

We want to increase the retention rate of game app users. Because the game has low satisfaction from the players and many of them uninstall the game.

Root Cause Analysis

- First gate in cookie cats is placed at a time that users are less interested in
- No discount items on app purchases
- No variation in game
- Segmentation campaign that is not right on target
- Users choose to play games that were previously played frequently
- Unattractive UI/UX
- Users don't have enough time to play games
- There is no reward when reaching a certain level in playing the game

Problem Statement

- Application users think that making purchases in game application is just for fun money and not profitable
- Customer do not want to pay extra fee
- Users do not like the game until the end of the level

Proposed Solution

- Changing the Open Gate to a higher level of 40

Key Metrics

Increase the retention rate of users game application

Population

Users who have been using cookie cat at least in the retention 1

Business Hypothesis

Changing the gate from gate 30 to gate 40 will increase the application retention rate

Hypothesis

H0: Changing Gate to 40 doesn't have a significant impact on the retention rate within one week

H1: Changing the Gate to 40 has a significant impact on the retention rate within one week

Experiment Groups and Period

Segment	Players
Control	19993
Treatment	19993

Control : variance at Gate 30

Experiment : variance at Gate 40

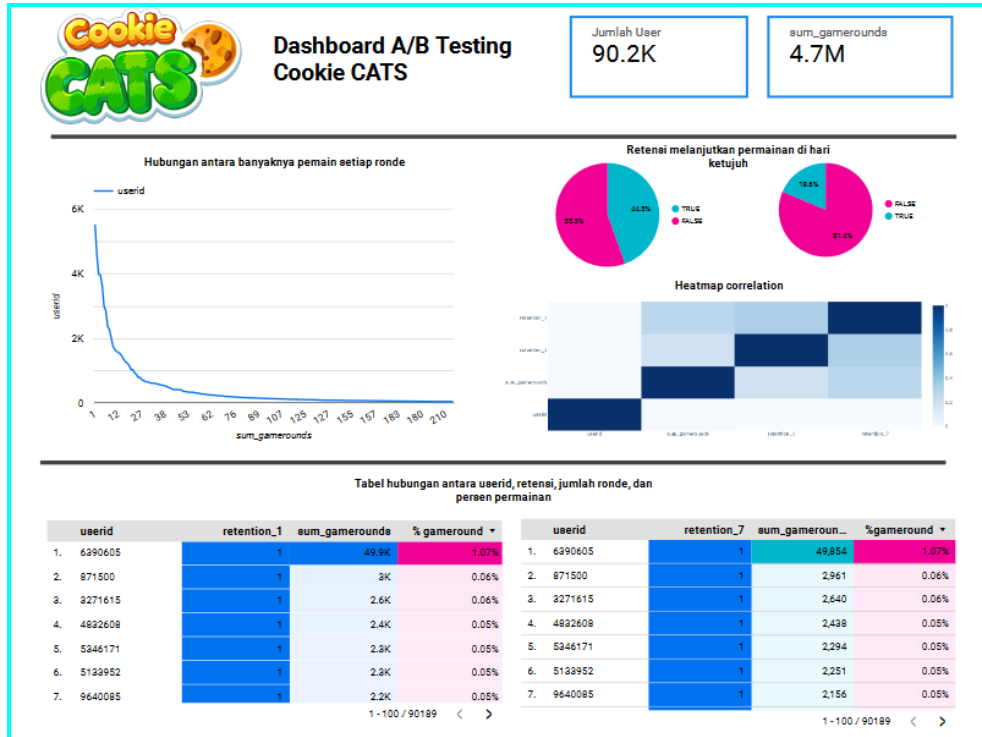
Experiment Monitoring

Experiment Mentoring used to create a dashboard with real-time data that has been recorded by system using Google Data Studio (GDS). These are some information come from monitoring result:

1. The relationship between the number of players and the game round, where the higher the interest in playing the smaller the round, which is indicated by the fewer number of players.
2. The number of players who use the application to play within a day after installing the application. Where on the first day playing was 44.5% of the total data, while on the seventh day it shrank back to only 18.6%.
3. Correlation information is obtained between the total column of data, where retention 7 has a correlation with the number of rounds.
4. Detailed data information is presented in tabular form.

Detailed dashboard can be access by this link:

<https://datastudio.google.com/reporting/28e72e60-a638-4a6d-b160-7d9a26aa618c>



Pre-processing Data

Prior to the dataset analysis, preprocessing was carried out to remove outliers, and to take only segments that would not cause bias. The preprocessing stages carried out include the following:

1. Checking Null Values

From the results of checking there is no data with null or empty values.

2. Remove Outlier from data

After checking process using visualization with box plot and bar plot results that there is outlier (Figure 1) in the sum round column. Therefore, one record in the form of outlier with values near 50.000 needs to be remove. The final result of data distribution after the outlier removing process can be seen on (Figure 2).

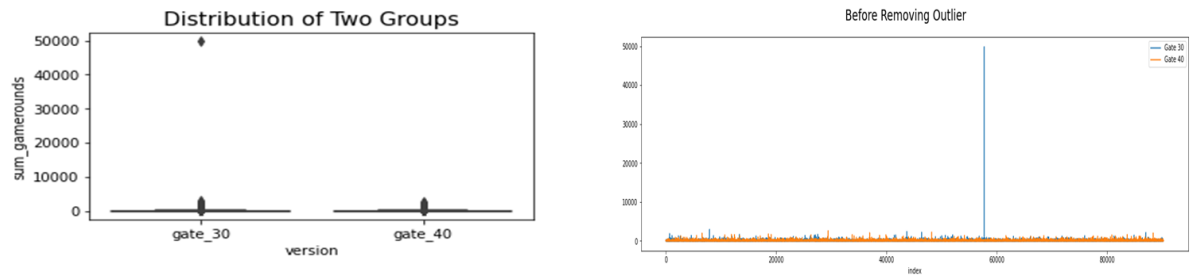


Fig 1. Dataset Distribution Before Delete Outlier

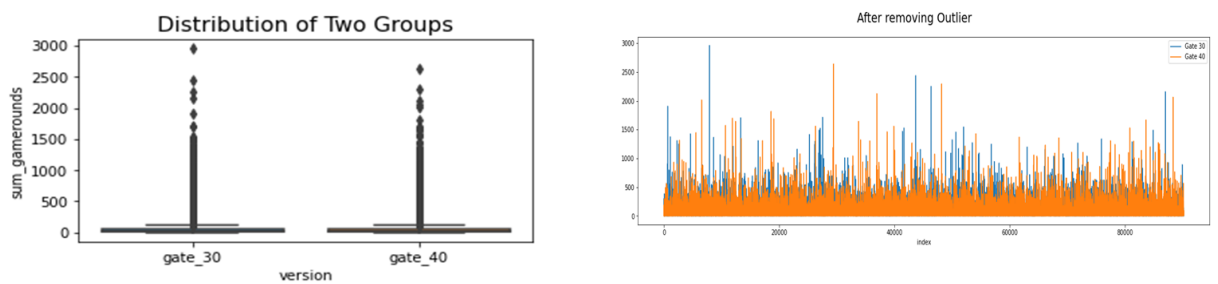


Fig 2. Dataset Distribution After Delete Outlier

3. Removing sum_gameround with 0 values

After observing the data, it founds that there are sum_gameround with 0 values as many as 3994 data. Furthermore, Can be conclude that players didn't play the game at all. Moreover, the comparison between players and sum_gameround add up from 0 and more than 0 can be seen in Figure 3.

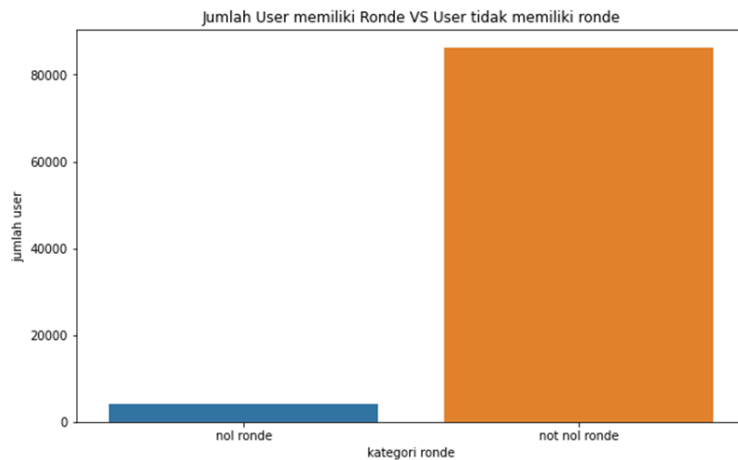


Fig 3. Number of users having rounds vs not having rounds

4. Removing players who have false value on retention 1

After we looked at the data, it was found that there were players who had a false value on retention one. Where in our opinion if this value is maintained there will be a bias towards the analysis carried out. So we decided to take data that has a value of true on retention 1 so we can see if the player who played on the first day will play the game again on day 7. That way we can assume if the player continues until day 7 then the player is interested in our game and become a loyal player in this game. After deleting player data that has a value of False on retention 1, the distribution of the data will look like in Figure 4.

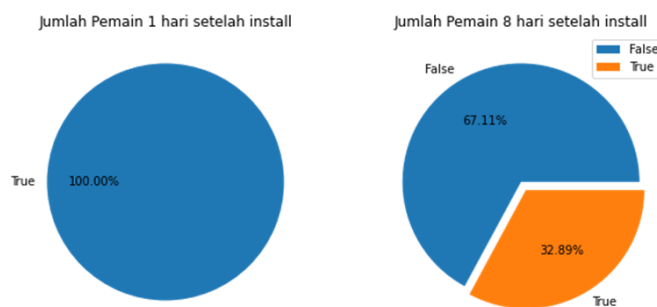


Fig 4. Final Data Distribution After Preprocessing

Analysis

Descriptive and inferential analysis based on experiment

1. Descriptive Analysis

a. Description of the Rounds on each Gate

Seen from the overall data, if the game is changed to Gate 40, the average number of rounds is more than the average game round with Gate 30, with a difference of 1 (Figure 5). In more detail, the seventh retention had a higher average gameround when compared to the first day retention at both Gate 30 and Gate 40. However, on the seventh day of Retention Gate 40 the number of people not playing games also increased (Figures 6 and 7).

	median	mean
version		
gate_30	48.0	94.605312
gate_40	50.0	95.657580

Fig 5. Average sum of game round in each Gate

		mean
version	retention_1	
gate_30	1	94.605312
gate_40	1	95.657580

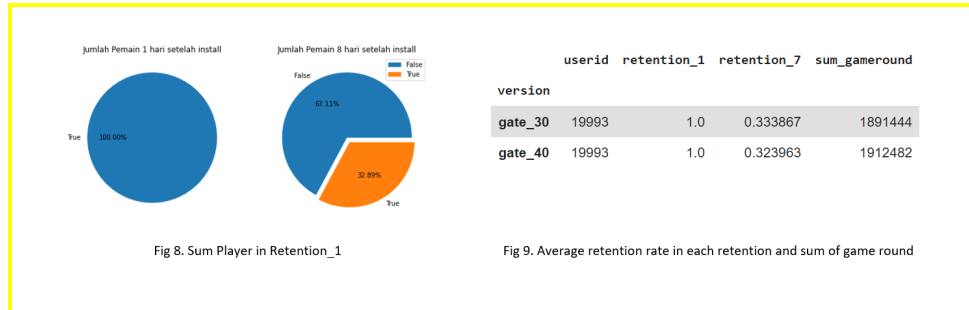
Fig 6. Average sum of game round in first retention

		mean
version	retention_7	
gate_30	0	49.843745
	1	183.913858
gate_40	0	50.221885
	1	190.471360

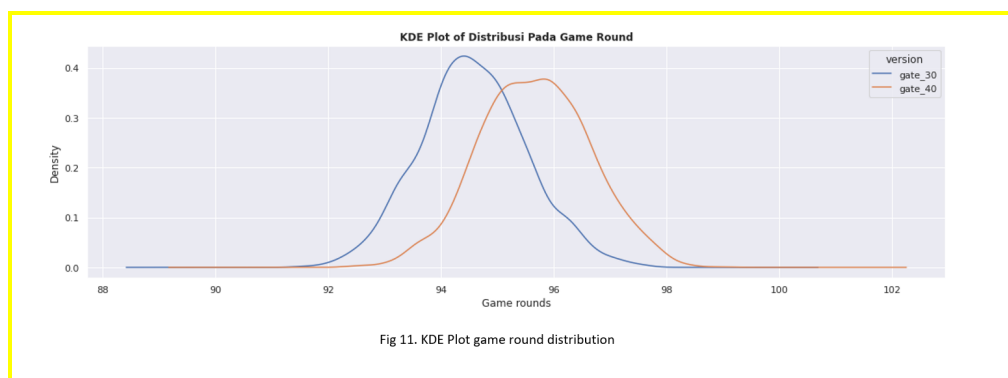
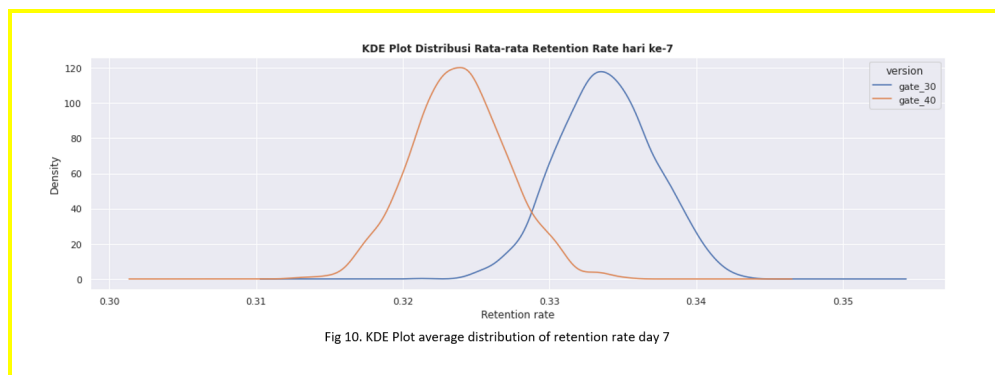
Fig 7. Average Sum of Game Round in seventh retention

b. Average Retention and Number of Rounds

Based on data analysis, the average number of users who return to play on retention 7 has decreased by 67% (Figure 8). In detail, Gate 40 experienced a decrease in the average number of players compared to the average Gate 30 player with a comparison of 33:32 or it can also be said to have a difference of 1. However, if viewed from the side of the number of rounds, Gate 40 is superior by 21,038 rounds (Figure 9).

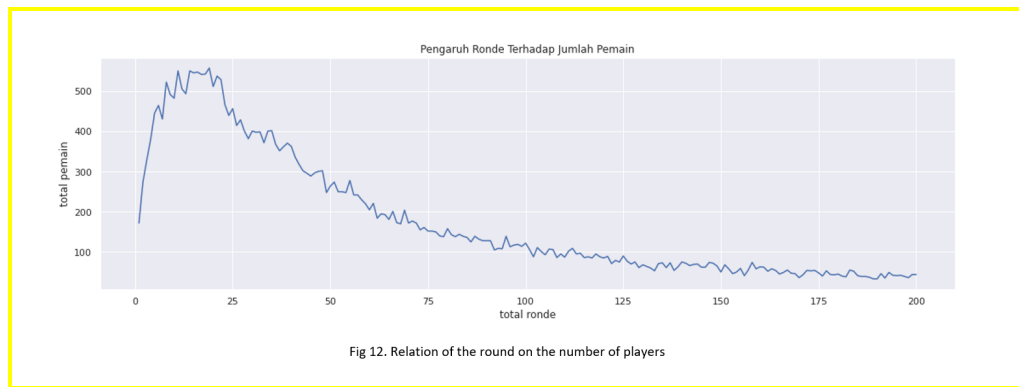


Based on the data visualization using KDE plots to find the distribution of the average retention rate in the seventh retention, it was found that Gate 40 has a smaller average retention rate than Gate 30. This can be seen in detail in (Figure 10). In addition, information about the number of rounds in (Figure 11) using the KDE plot visualization to determine the distribution of the number of game rounds was obtained, the fact that the number of game rounds at gate 40 was higher, this can be assumed because gate 40 has a higher level which causes the number of rounds is also in line.



c. Effect of rounds on player interest

Based on the experiments and the line plot shown in Figure 12 (Figure 12). It was found that the rounds in the game have an influence on the interest of the players. Where the higher the interest in the game, the fewer players this can be seen from the number of players that continues to decrease. The highest number of players is only between rounds 5 to 25. This may be due to a number of reasons, such as the user's lack of interest in continuing the game, poor UI/UX, game challenges that get harder as the number of rounds increases, and many more.



1. Inferential Analysis

a. Correlation

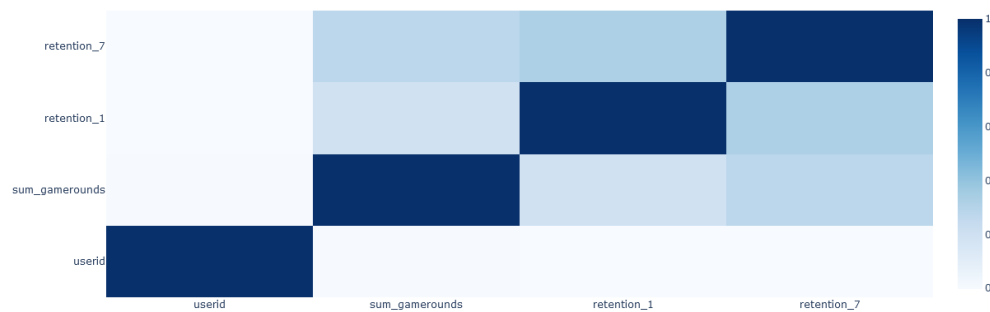


Figure 13

Based on the heatmap above(Figure 13), it can be seen that there is a correlation between retention 1 and retention 7 with a correlation value is 0.327, the correlation between retention 1 and sum_gameround is 0.197 and the correlation between retention 7 and sum_gameround is 0.279

b. Linear Regression

	Coefficient
sum_gameround	0.0
userid	0.0

Figure 14

Coefficient in retention 1

	Coefficient
sum_gameround	1.610838e-03
userid	-8.140898e-10

Figure 15

Coefficient in Retention 7

This means that sum_gameround and userid have not effect on retention _1 (Figure 14). For retention_7 means that "sum_gameround" increase of 1.610838 and there is a decrease of 8.140898 user in userid. We can see that "userid" have a little effect on the retention_7 (Figure 15).

• Evaluating

The final step is to evaluate the performance of the linear regression algorithm. We'll do this by finding the values for MAE, MSE and RMSE. This the output:

Figure 16

Retention 1

```
Mean Absolute Error: 0.0
Mean Squared Error: 0.0
Root Mean Squared Error: 0.0
```

Figure 17
Retention 7

```
Mean Absolute Error: 0.3563012189986435
Mean Squared Error: 0.1689990269002515
Root Mean Squared Error: 0.41109491227726414
```

You can see in retention 1 the MAE, MSE and RMSE have a zero value because in preprocessing it has been determined that the retention 1 value used is only True (Figure 16). While the value of root mean squared error (RMSE) in retention 4 is 0.411 (Figure 17). This means that our algorithm was accurate and can make reasonably good predictions.

c. Chi-Square

Based on statistical calculations using Chi-Square on control and treatment variables on retention 7, the Chi-square value is 4.4, the p-value is 0.03. Where from the p-value obtained, it can be concluded that there is a significant effect if the gate is changed to 40, which is 5%. From these calculations, the final statistical analysis is to reject H0 and accept H1 or in another language, namely, H0 is not the same as h1. Details of the calculation results can be seen in Figure 18.

```
Chi-square: 4.4418217519009815
p-value: 0.035068807842018185
Degree of freedom: 1
[[13417.  6576.]
 [13417.  6576.]]
Significant at 5% level of significance
```

Fig 18. calculation results using chi square

Action Plan

Based on experiment we can conclude that:

1. The change on Gate from Gate 30 to Gate 40 gives a significant change by 5%
2. The total round on Gate 40 is higher than Gate 40 because of the total linear level
3. On Gate 40, the average number of retention rates is lower if compared with Gate 30. Therefore, it is concluded that users will be interested to Gate 30 by 5% compared with Gate 40
4. On the graphic, it is stated that the higher number of total rounds means higher number of Gates. This will also reducing the number of users.

Based on our conclusion, we suggest some Action Plans such as:

1. Do not change the gate position from Gate 30 to Gate 40
2. Observe and find other variables that may give effect to increase retention rate significantly
3. Carry out further experiments using different variables
4. Add more complex data such as bigger retention to daily or weekly, so the analysis can be done more accurate