



A/B TESTING HYPOTHESIS TESTING

dev.abdullah.khaled@gmail.com
+20 1557504902



Project Description from DataCamp

1

- 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. We're not kidding!
- 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 A/B test 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.



Data Description from Aurelia Sui's notebook

2

The data is from 90,189 players that installed the game while the AB-test was running. The variables are:

- **userid** - a unique number that identifies each player.
- **version** - whether the player was put in the control group (gate_30 - a gate at level 30) or the test group (gate_40 - a gate at level 40).
- **sum_gamerounds** - the number of game rounds played by the player during the first week after installation
- **retention_1** - did the player come back and play 1 day after installing?
- **retention_7** - did the player come back and play 7 days after installing?

When a player installed the game, he or she was randomly assigned to either gate_30 or gate_40.



AB Testing Process

3

1. Understanding business problem & data
2. Detect and resolve problems in the data (Missing Value, Outliers, Unexpected Value)
3. Look summary stats and plots
4. Apply hypothesis testing and check assumptions:
 - Check Normality & Homogeneity
 - Apply tests (Shapiro, Levene Test, T-Test, Welch Test, Mann Whitney U Test)
5. Evaluate the results
6. Make inferences
7. Recommend business decision to your customer/director/ceo etc.



The Dataset

4

	userid	version	sum_gamerounds	retention_1	retention_7
0	116	gate_30	3	False	False
1	337	gate_30	38	True	False
2	377	gate_40	165	True	False
3	483	gate_40	1	False	False
4	488	gate_40	179	True	True

SUMMARY STATS

5

sum_gamerounds

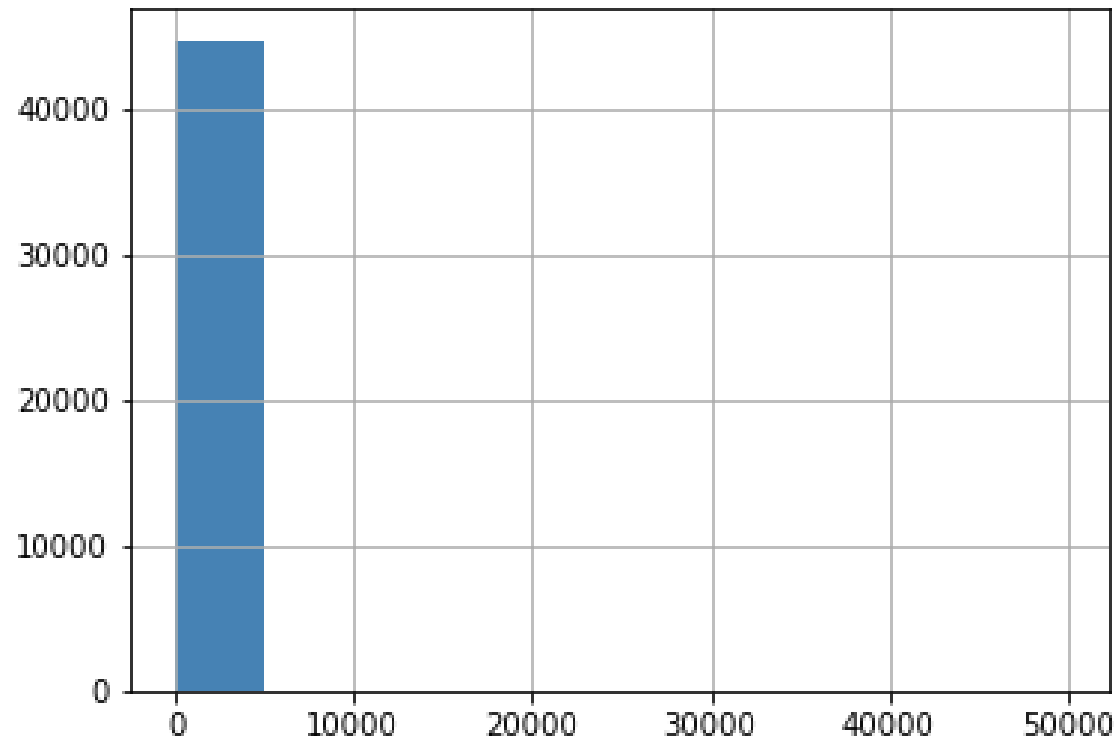
count	mean	std	min	1%	5%	10%	20%	50%	80%	90%	95%	99%	max
90189.0000	51.8725	195.0509	0.0000	0.0000	1.0000	1.0000	3.0000	16.0000	67.0000	134.0000	221.0000	493.0000	49854.0000

A/B Groups & Target Summary Stats

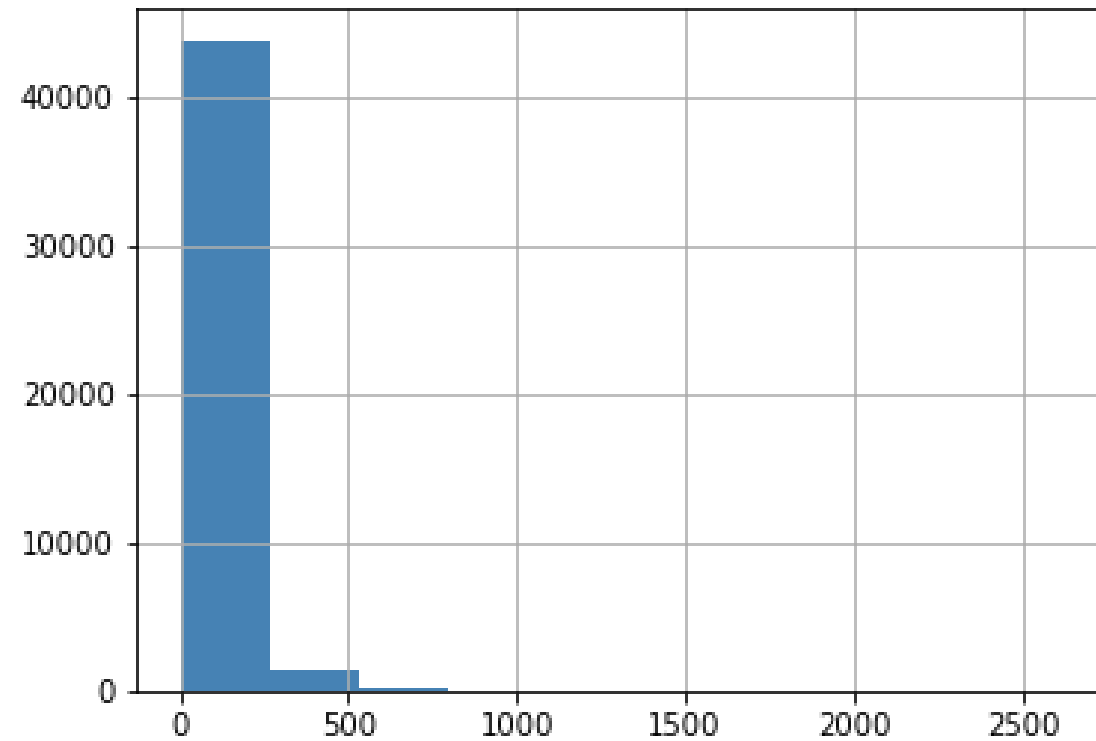
	count	median	mean	std	max
version					
gate_30	44700	17	52.4563	256.7164	49854
gate_40	45489	16	51.2988	103.2944	2640

Before Removing The Extreme Value

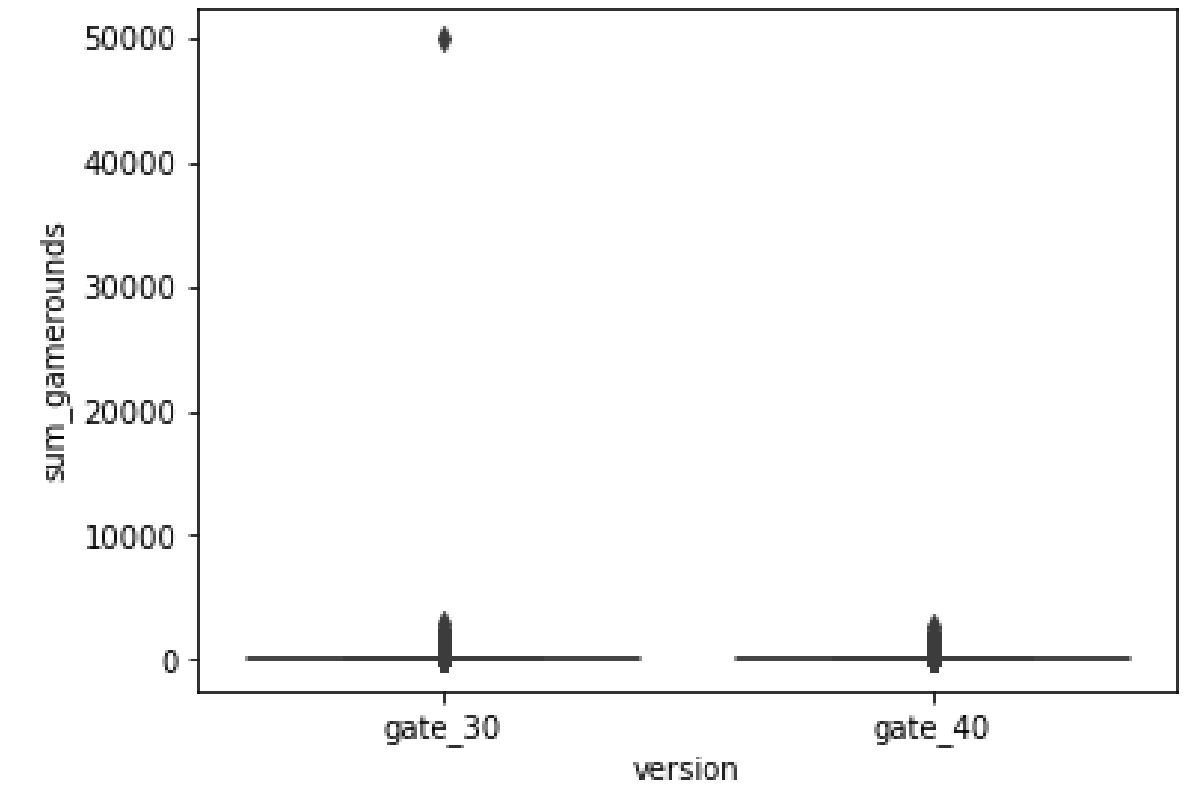
Distribution of Gate 30 (A)



Distribution of Gate 40 (B)



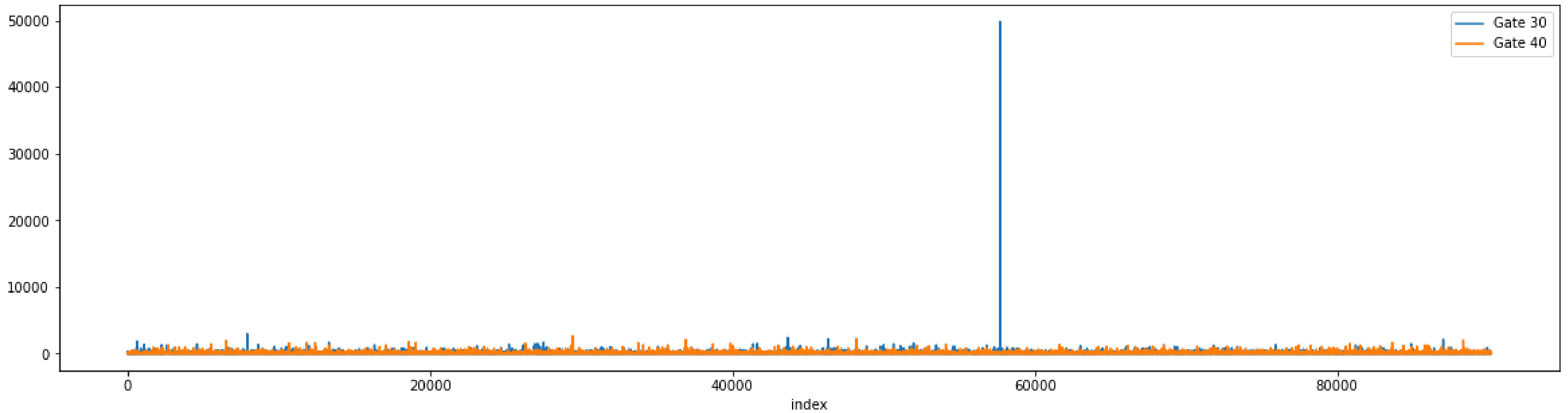
Distribution of Two Groups



Before Removing The Extreme Value

SUMMARY STATS

5



Before Removing The Extreme Value

sum_gamerounds

count	mean	std	min	1%	5%	10%	20%	50%	80%	90%	95%	99%	max
90188.0000	51.3203	102.6827	0.0000	0.0000	1.0000	1.0000	3.0000	16.0000	67.0000	134.0000	221.0000	493.0000	2961.0000

A/B Groups & Target Summary Stats

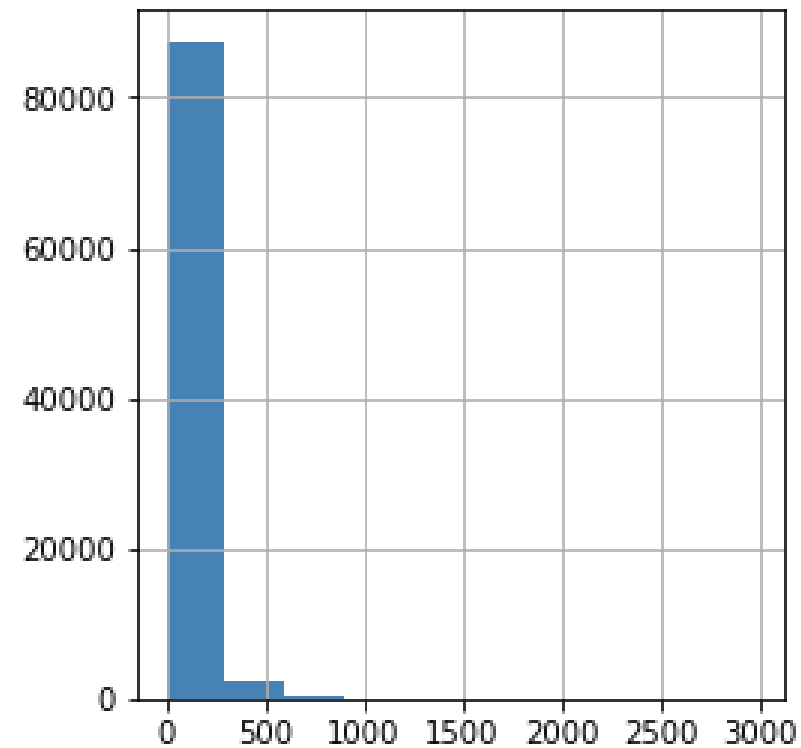
	count	median	mean	std	max
version					
gate_30	44699	17	51.3421	102.0576	2961
gate_40	45489	16	51.2988	103.2944	2640

After Removing The Extreme Value

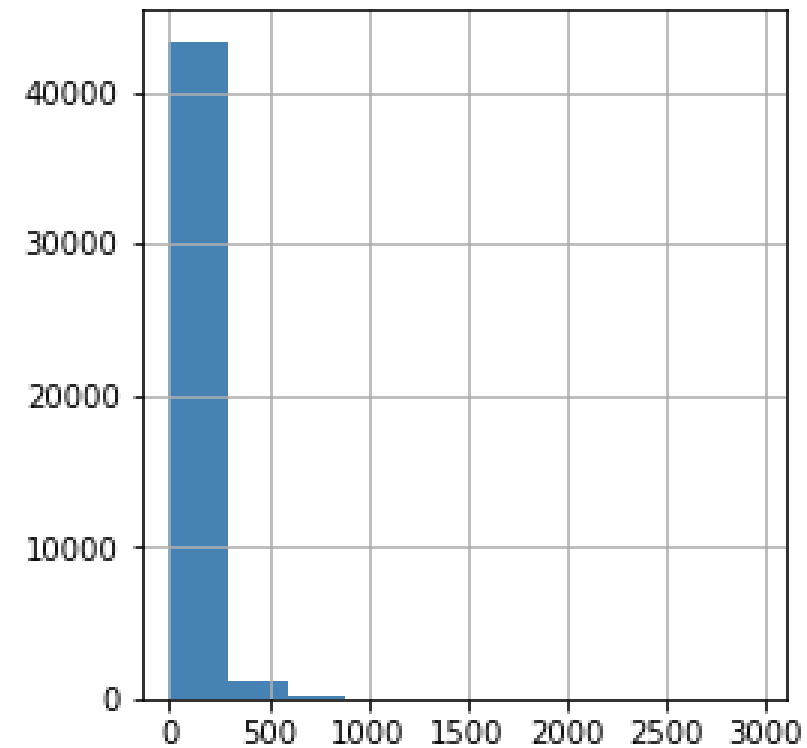
SUMMARY STATS

5

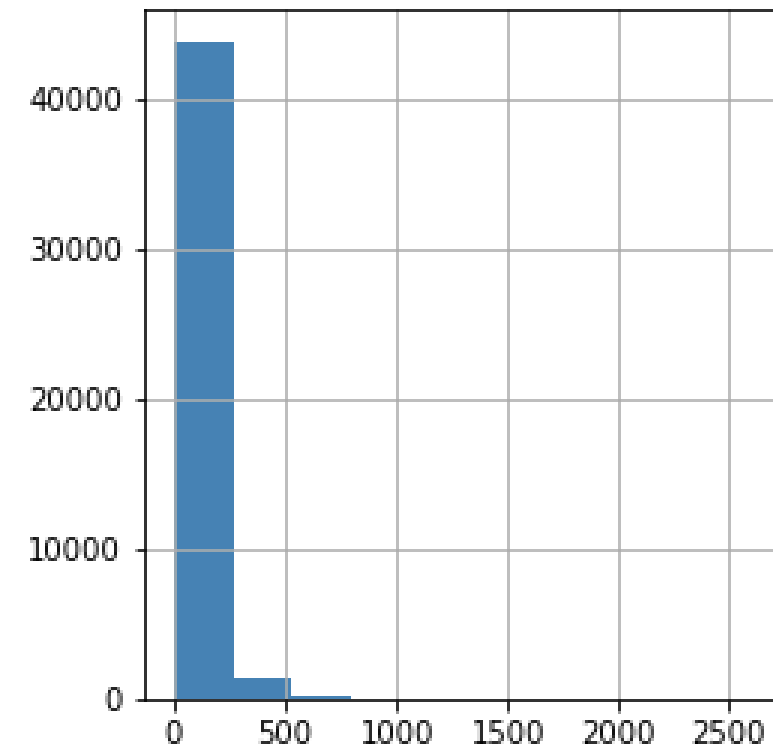
Distribution of Total Game Rounds



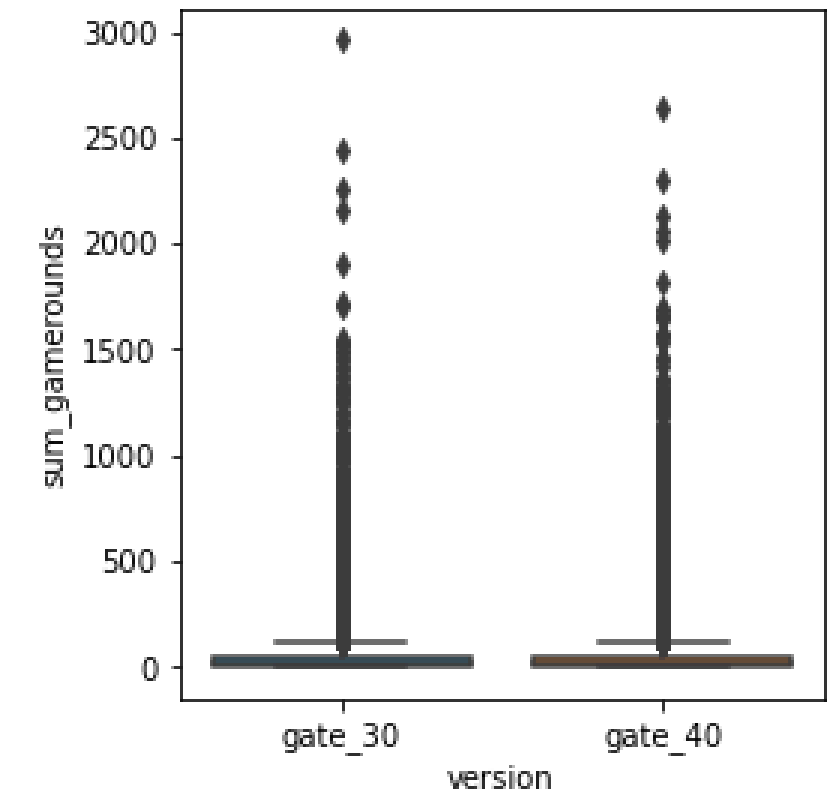
Distribution of Gate 30 (A)



Distribution of Gate 40 (B)



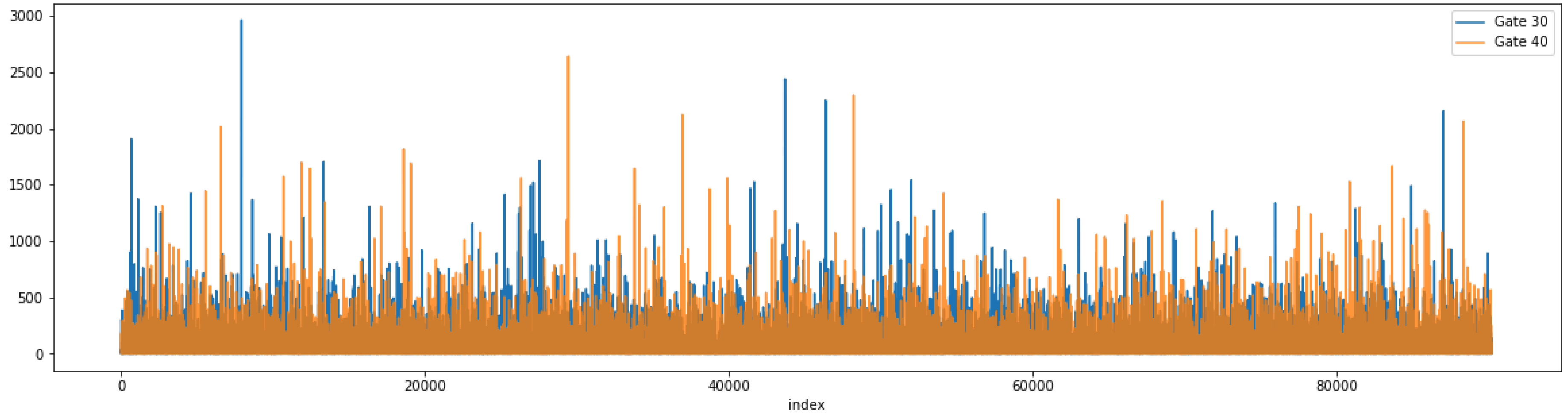
Distribution of Two Groups



After Removing The Extreme Value

SUMMARY STATS

5



After Removing The Extreme Value

The users installed the game but 3994 users never played the game! Some reasons might explain this situation:

They have no free time to play game.

Users might prefer to play other games or they play other games already.

Some users don't like the app etc.

The number of users decreases as the levels progress

Most of users played the game at early stage and they didn't progress.

Tactile Entertainment should learn why users churn playing the game.

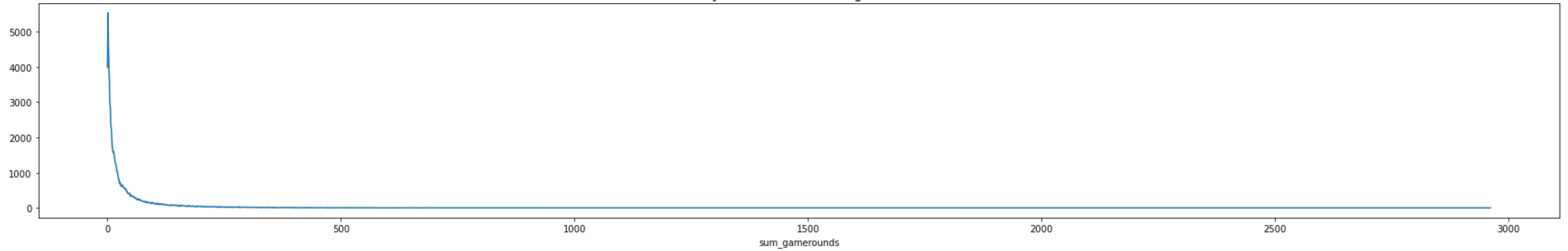
Doing research and collecting data about the game and users would help to understand user churn

SOME DETAILS

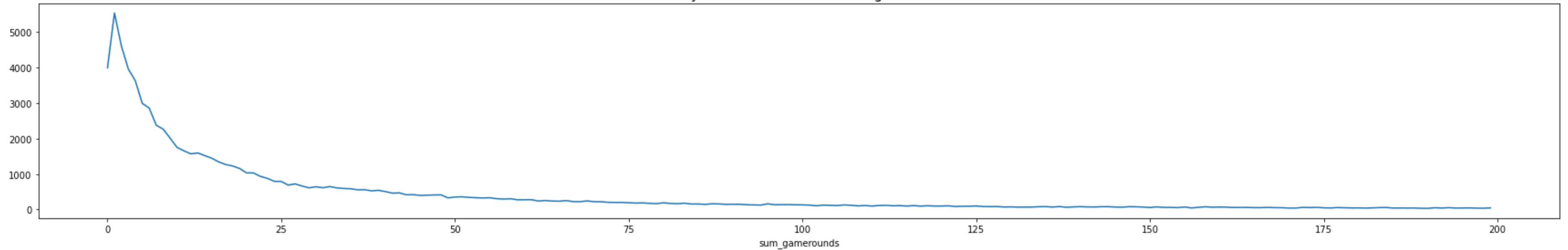
6

The number of users in the game rounds played

How many users are there all game rounds?



How many users are there first 200 game rounds?



Retention variables gives us player retention details.

- **retention_1** - did the player come back and play 1 day after installing?
- **retention_7** - did the player come back and play 7 days after installing?
- Also players tend not to play the game! There are many players who quit the game.
- **55 percent** of the players didn't play the game 1 day after installing
- **81 percent** of the players didn't play the game 7 day after installing

	RET1_COUNT	RET7_COUNT	RET1_RATIO	RET7_RATIO
False	50035	73408	0.5548	0.8139
True	40153	16780	0.4452	0.1861

SOME DETAILS

6

Looking at the summary statistics of retention variables by version and comparing with `sum_gamerounds`, there are similarities between groups. However, it will be more helpful to see if there is a statistically significant difference.

		count	median	mean	std	max
version	retention_1					
gate_30	False	24665	6	16.3591	36.5284	1072
	True	20034	48	94.4117	135.0377	2961
gate_40	False	25370	6	16.3404	35.9258	1241
	True	20119	49	95.3812	137.8873	2640

		count	median	mean	std	max
version	retention_7					
gate_30	False	36198	11	25.7965	43.3162	981
	True	8501	105	160.1175	179.3586	2961
gate_40	False	37210	11	25.8564	44.4061	2640
	True	8279	111	165.6498	183.7925	2294

SOME DETAILS

6

Similar results are seen when the number of users who came and did not come 1 day and 7 days after the game was installing. Approximately 12.000 users among the total users played the game both 1 day and 7 days after installing the game. 14% of the total users include people who will continue the game in the future.

		count	median	mean	std	max
version	Retention					
gate_30	0	38023	12	28.0703	48.0175	1072
	1	6676	127	183.8863	189.6264	2961
gate_40	0	38983	12	28.1034	48.9278	2640
	1	6506	133	190.2824	194.2201	2294

SOME DETAILS

6

When the retention variables are combined and the two groups are compared, the summary statistics are similar here as well.

	version	NewRetention	count	median	mean	std	max
0	gate_30	False-False	22840	6	11.8197	21.6426	981
1	gate_30	False-True	1825	43	73.1693	93.2223	1072
2	gate_30	True-False	13358	33	49.6945	58.1254	918
3	gate_30	True-True	6676	127	183.8863	189.6264	2961
4	gate_40	False-False	23597	6	11.9133	20.9010	547
5	gate_40	False-True	1773	47	75.2611	94.4780	1241
6	gate_40	True-False	13613	32	50.0255	60.9246	2640
7	gate_40	True-True	6506	133	190.2824	194.2201	2294

Assumptions:

- Check normality
- If Normal Distribution, check homogeneity

Steps:

- Split & Define Control Group & Test Group
- Apply Shapiro Test for normality
- If parametric apply Levene Test for homogeneity of variances
- If Parametric + homogeneity of variances apply T-Test
- If Parametric - homogeneity of variances apply Welch Test
- If Non-parametric apply Mann Whitney U Test directly

Define A/B groups

	userid	version	sum_gamerounds	retention_1	retention_7	Retention	NewRetention
0	116	A	3	False	False	0	False-False
1	337	A	38	True	False	0	True-False
2	377	B	165	True	False	0	True-False
3	483	B	1	False	False	0	False-False
4	488	B	179	True	True	1	True-True

The Result

```
# A/B Testing Hypothesis
```

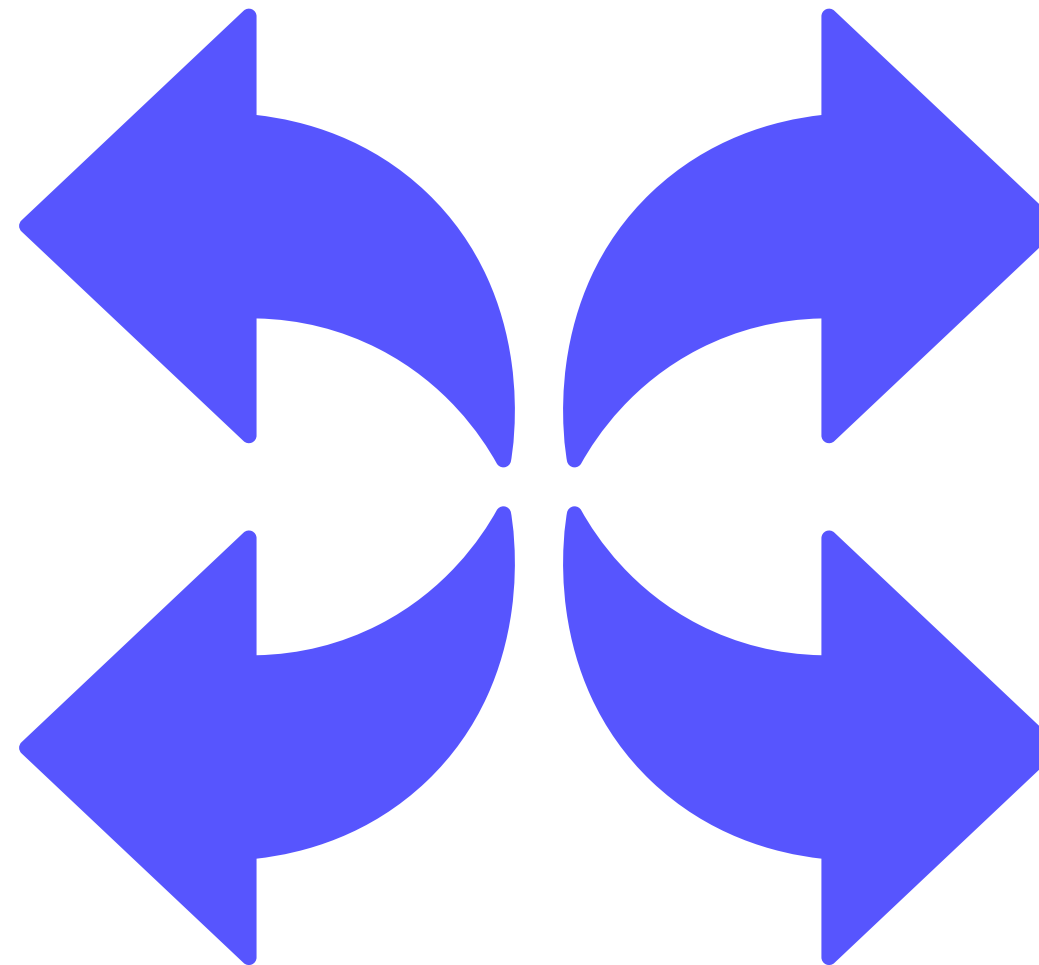
```
H0: A == B
```

```
H1: A != B
```

	Test Type	AB Hypothesis	p-value	Comment
0	Non-Parametric	Reject H0	0.0254	A/B groups are not similar!

Firstly, we investigated relationships and structures in the data. There was no missing value problem but was one outlier problem in the data. Summary stats and plots help us to understand the data and problem.

Before A/B Testing, we shared some details about game, players, problems and suggestion to our customer/director/ceo etc.



After applying A/B Testing, the analysis result gives us some important information. Shapiro Testing rejected H_0 for Normality assumption. Therefore we needed to apply a Non-parametric test as called Mann Whitney U to compare two groups. As a result, Mann Whitney U Testing rejected H_0 hypothesis and we learned A/B groups are not similar!

Briefly, There are statistically significant difference between two groups about moving first gate from level 30 to level 40 for game rounds.

Which level has more advantages in terms of player retention?

- 1-day and 7-day average retention are higher when the gate is at level 30 than when it is at level 40.

```
(version
A    0.4482
B    0.4423
Name: retention_1, dtype: float64,
version
A    0.1902
B    0.1820
Name: retention_7, dtype: float64)
```



THANK YOU

dev.abdullah.khaled@gmail.com
+20 1557504902

