KM DATA 4 - GROUP 5

COOKIE CATS EXPERIMENT TESTING

HOMEWORK A (Odd Group)

GOAL/OBJECTIVE

Currently, the first gate was at level 30. The company wants to know the effect of moving the first gate from level 30 to level 40. Would the change affect player retention and game rounds?

GENERAL PROBLEM

Changes:
Moved the gate from level 30 to level 40 in Cookie Cats

Goal:
To increase retention rate and conversion rate (buy rate)

PROBLEM STATEMENT

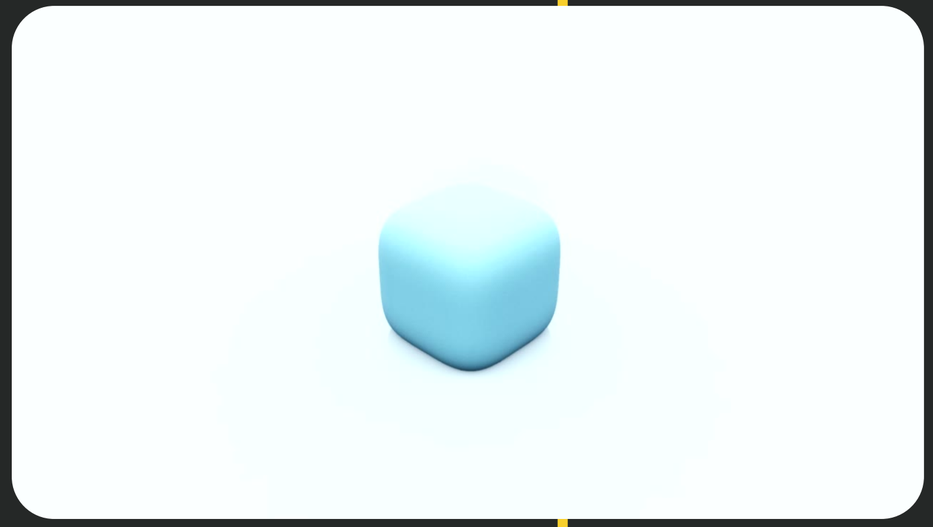
The Dataset is about the mobile game Cookie Cats pops. At first, players download the game for free. As players progress through the game, they would encounter gates that ask them to wait some amount of time to enter the next game or make purchases to avoid waiting. As a result, the timing of placing gates is an important decision to make.

PROPOSED SOLUTION

1. **Understand** the dataset
2. EDA and data **visualization**
3. **Hypothesis** testing - bootstrapping
4. **Result** and business **recommendation**

KEY METRICS

To improve retention rate as our goal, we need to **choose retention rate** for matrics. Let retention rate for gate_30 and gate_40 be P30 and P40.



POPULATION

90,189 players that installed the game **while** the AB-test was running

Hypothesis

- H0: There is no difference between the gate 30 and gate 40 versions in terms of player retention and game rounds.
- H1: There is a difference between the gate 30 and gate 40 versions in terms of player retention and game rounds.

Business Hypothesis

There is an effect of moving the first gate from level 30 to level 40 on player retention and game rounds.

Experiment Group & Period

The dataset contains 90189 player data. There are two versions: gate_30, gate_40, total games played by the player, and if the player logs back in after day1 and day7.

Experiment Monitoring

Monitoring the difference in player retention and game rounds at gates 30 and 40.

Explore Dataset

Let's take a look at our dataset

	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

We have 90,189 rows and 5 columns :

- userid : a unique number to identify a player
- version : When a player installed the game, they were randomly assigned to either gate_30 or gate_40
- sum_gamerounds : the number of game rounds played by the player during the first 14 days after install
- retention_1 : did the player come back and play 1 day after installing?
- retention_7 : did the player come back and play 7 day after installing?

Explore Dataset

Let's calculate total number of players, 1-day and 7-days retention and sum_gamerounds for each A group (gate_30) and B group (gate_40)

version	userid	retention_1	retention_7	sum_gamerounds
gate_30	44700	0.448188	0.190201	2344795
gate_40	45489	0.442283	0.182000	2333530

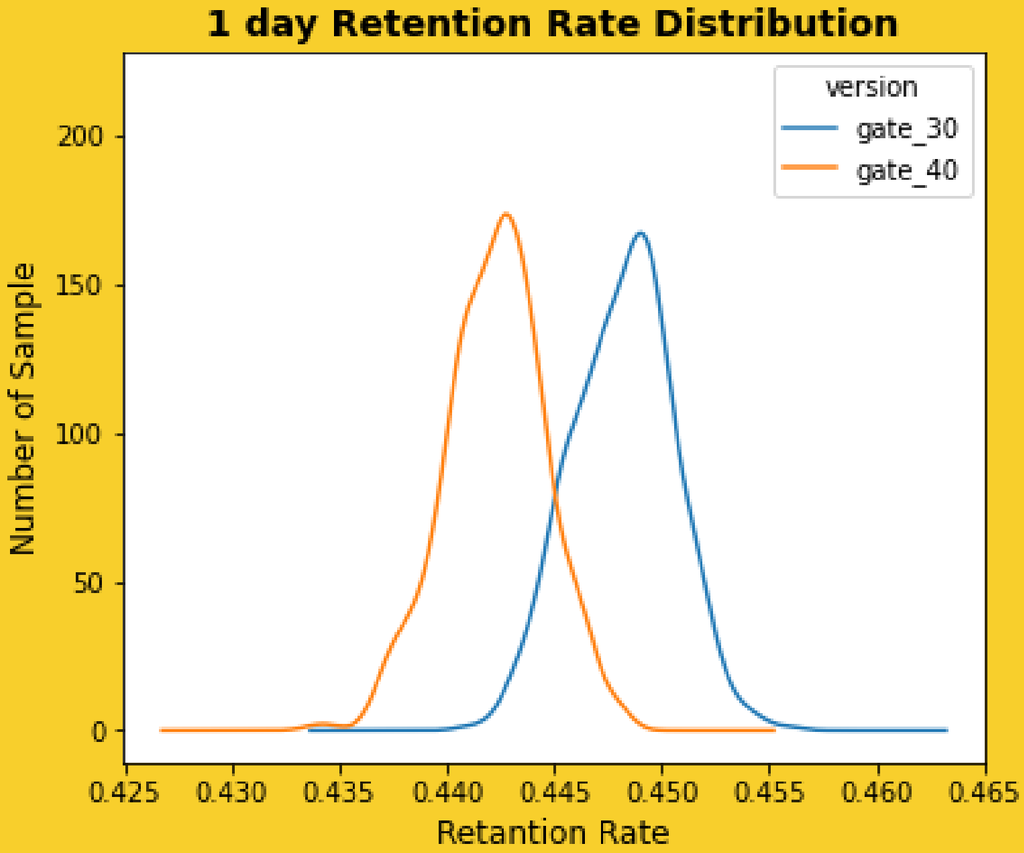
From the result above we can see that ：

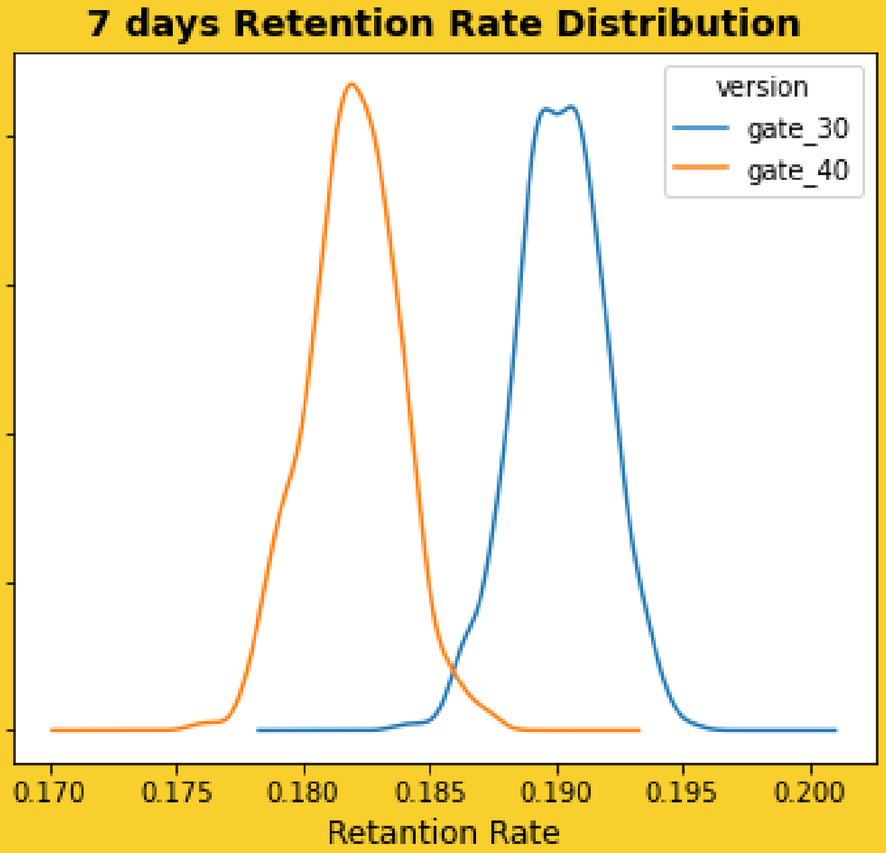
- There was a slight decrease in 1-day retention when the gate was moved to level 40 from 44.8% when it was at level 30 to 44.2%

- Again,a decrease in 7-day retention when the gate was moved to level 40 from 19% when it was at level 30 to 18.2% .

Explore Dataset

The bootstrap method is a resampling technique used to estimate statistics on a population by sampling a dataset with replacement. Here's the result after creating 500 sample using bootstrap method :





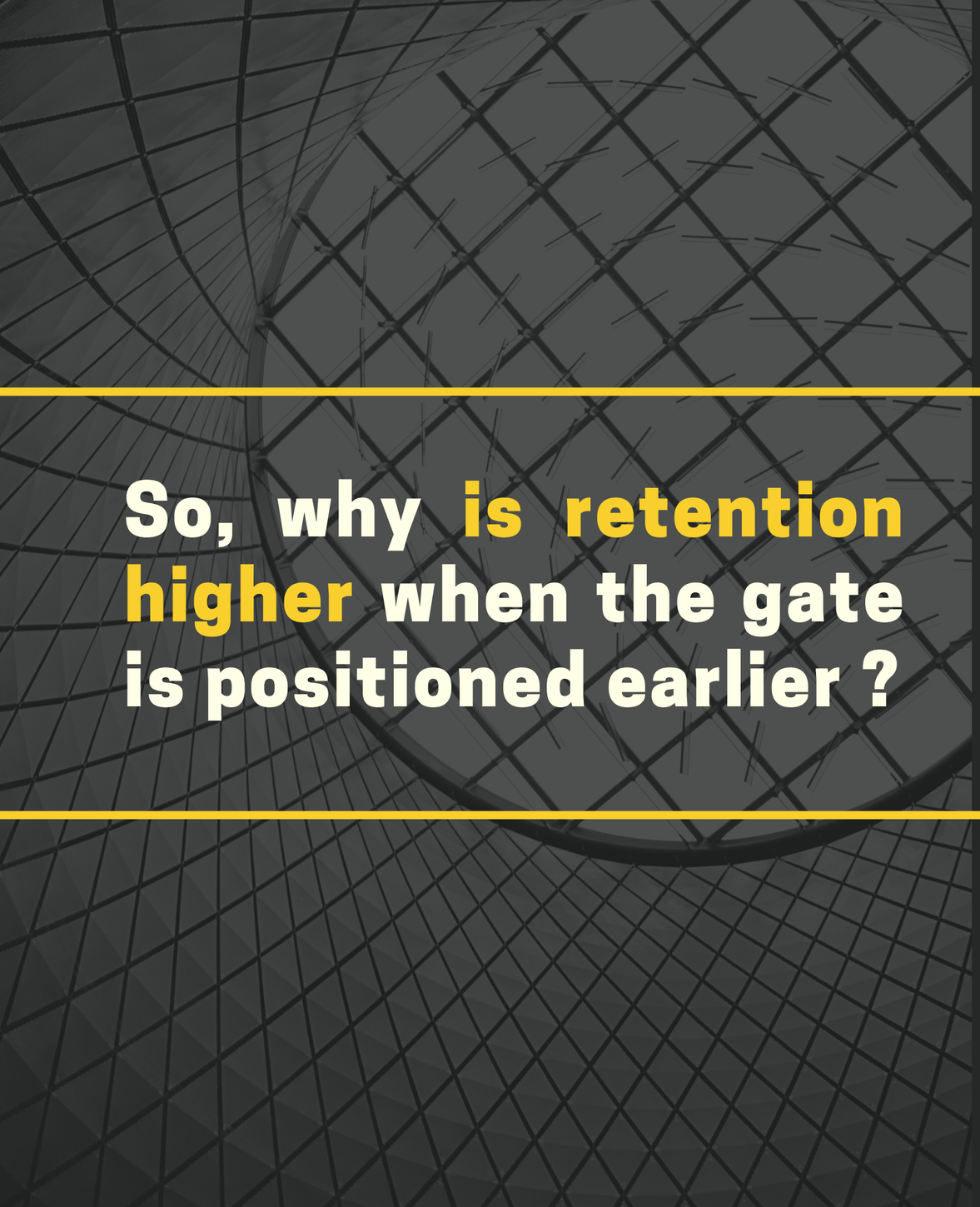
Explore Dataset

Just eyeballing the plot before, we can see that there seems to be some evidence of a difference. So, let's find out the probability using those difference

```
# Adding a column with the % difference between the two A/B groups
boot_1d['diff'] = ((boot_1d['gate_30'] - boot_1d['gate_40']) / boot_1d['gate_40'] * 100)
boot_7d['diff'] = ((boot_7d['gate_30'] - boot_7d['gate_40']) / boot_7d['gate_40'] * 100)

# Calculating the probability that 1-day retention is greater when the gate is at level 30
prob_1 = (boot_1d['diff']>0).sum()/len(boot_1d['diff'])

# Calculating the probability that 7-days retention is greater when the gate is at level 30
prob_7 = (boot_7d['diff']>0).sum()/len(boot_7d['diff'])

# Pretty printing the probability
print(f"The probability that 1-day retention is greater when the gate is at level 30: {round(prob_1,2)*100}% \
      \nThe probability that 7-days retention is greater when the gate is at level 30: {(prob_7)*100}% ")
```

```
The probability that 1-day retention is greater when the gate is at level 30: 96.0%
The probability that 7-days retention is greater when the gate is at level 30: 100.0%
```

The bootstrap result tells us that there is strong evidence probability that both 1 and 7-day retention is higher when the gate is at level 30 than when it is at level 40.



So, why is retention higher when the gate is positioned earlier ?

LEVEL 30 BETTER THAN LEVEL 40

Hedonic adaptation is the tendency for people to get less and less enjoyment out of a fun activity over time if that activity is undertaken continuously.

Why? There will be fewer players who manage to reach a considerable distance, and they tend to quit the game because they get bored when the gate is moved to level 40.

DECISION

Our decision is **not** to move the gate from level 30 to level 40 If we want to maintain high retention.
Fine retention of 1 day or 7 days.

RESULT ?

Bootstrapping results tell us that there is strong evidence that 1-day and 7-day **retention is higher when the gate is at level 30** than when it is at level 40.

FUTHER STUDY...

We will be able to measure the difference in conversion rate (in-purchase rate) in the number of rounds played.

Then we can also see the level of retention that occurs.

STUDY CAN BE CONDUCTED

BEHIND THE RESULT:

Common sense is that The later the obstacle, the longer people are going to engage with the game. But this is not what the data tells us.