Mobile Game "Brutal Age" New User Data Analysis

Heather Huang

I. Background and Purpose

- Background
  "Brutal Age": Horde Invasion is an MMO strategy mobile app game. Players are able to build a powerful Horde by building outposts, expanding territory, and competing with players from all over the world together. In-Game you can purchase from $0.99 to $99.99 USD. Depending on what was purchased it could supplement your gameplay or fast track it. The dataset which is provided by Tap4fun includes new user's gaming information during the first 7 days.

- Purpose
  Our goal is to provide insights and actionable tips on advertising and promotions strategies by analyzing new users gaming information data.
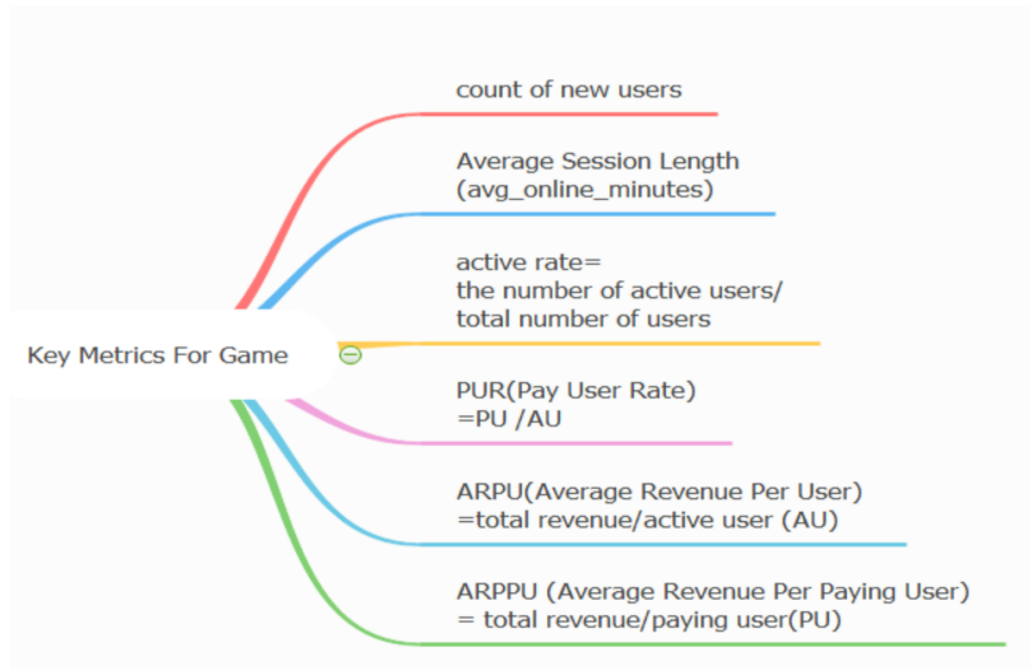
II. Method

- This report will use *MySQL*/*Navicat* to do analysis and use Excel to do data visualization.

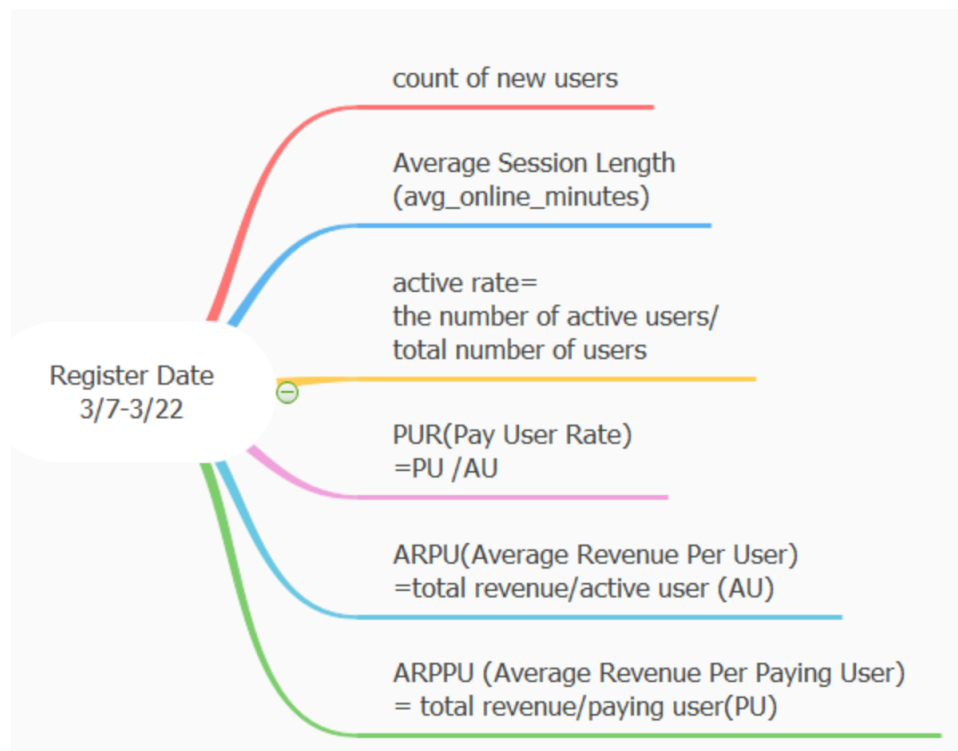- This report will answer the following questions:
  1. Does the game run well for new users?
  Track key metrics to give an overview of the development of 'Brutal Age'. Track these users' metrics include counts of new users, average session length, active user rate, PU (paying user rate), ARPU (Average revenue per user), ARPPU (Average revenue per paying user). By using the register date, paying users, non-paying users etc, to give an overview of the game's operation condition for new users:

Key Metrics For Game
- count of new users
- Average Session Length (avg_online_minutes)
- active rate= the number of active users/ total number of users
- PUR(Pay User Rate) =PU /AU
- ARPU(Average Revenue Per User) =total revenue/active user (AU)
- ARPPU (Average Revenue Per Paying User) = total revenue/paying user(PU)

2. Does different register dates affect these key metrics?
The data set includes 16 continuous days' data. By analyzing the number of new users per day, user payment per day, the active rate per day, we can figure out some special events and what effects did the specific event has on the game.

Register Date 3/7-3/22
- count of new users
- Average Session Length (avg_online_minutes)
- active rate= the number of active users/ total number of users
- PUR(Pay User Rate) =PU /AU
- ARPU(Average Revenue Per User) =total revenue/active user (AU)
- ARPPU (Average Revenue Per Paying User) = total revenue/paying user(PU)

III. Understand the dataset

There are total 828934 observations and 108 variables including user_id, register time, supplies (wood, stone, ivory, meat, magic, etc.), branch of the armed forces(infantry、 cavalry、 shaman), acceleration items(general, building, research, training、 treatment), building, research, PVP, PVE, avg_online_minutes、 pay price, pay count and so on. For this project we will only consider user information variables.
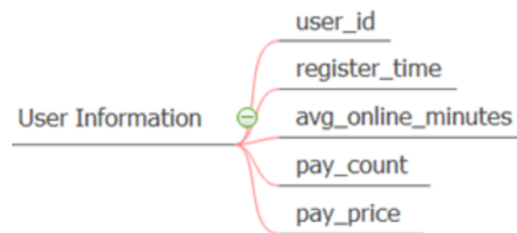
IV Data Preparation

- **Data Importing**

  Import data into MySQL database, add primary key, and format the columns' data type. pay_price decimal(50,2)， avg_online_minutes decimal(50,8)， register_time varchar(50)， change other columns' data type to int(11), tinyint(4) based on their value.

user_id	int	11	0	☑	🔑1
register_time	varchar	50	0	☐	
wood_add_value	int	11	0	☐	
wood_reduce_value	int	11	0	☐	
stone_add_value	int	11	0	☐	
stone_reduce_value	int	11	0	☐	
ivory_add_value	int	11	0	☐	
ivory_reduce_value	int	11	0	☐	
meat_add_value	int	11	0	☐	

pvp_win_count	int	11	0	☐
pve_battle_count	int	11	0	☐
pve_lanch_count	int	11	0	☐
pve_win_count	int	11	0	☐
avg_online_minutes	decimal	50	8	☐
pay_price	decimal	50	2	☐
pay_count	int	11	0	☐
time	timestamp	0	0	☑

▪ Data Cleaning and Data Transformation

Check if there's any missing data or duplicates. There are 828934 rows and
no missing values.

```
select count(user_id) from data;
select count(distinct user_id) from data;
select count(user_id) from `data` where user_id is null or user_id
='';
```

count(distinct user_id)
▸ 828934

count(user_id)
▸ 828934
0

- Convert the register_time column from string to timestamp. Then check the missing value for time. Here there is no missing value for time.

```
/*creat a new column to store the register_time in timestamp*/
alter table `data` add time TIMESTAMP not NULL;

/*update the new colomn with register_time*/
update data set time =
from_unixtime(round(unix_timestamp(register_time),0));

/*check if there's missing value*/
select count(time) from `data` where time is null union select
count(time) from data where time ='';
```

count(time)
0

V Data Analysis

Question 1: Does the game run well now?
Track key metrics to give an overview of the development of 'Brutal Age'.

- *Average Daily New Users: 51808*

```
/* average daily new user from 3/7 to 3/22*/
select avg(n) from (select count(distinct user_id) as n from data
group by date(time))as temp;
```

avg(n)
51808.3750

- *Average Session Length:11.74 minutes*
  From the result, we can tell 75% users' average online time are less than 5 minutes. Most users are inactive users.

```
/*Median,mean, Q1,Q3, min, max of average session length for all new
users during 3/7-3/22*/
select round(avg(avg_online_minutes),2) as avg,
round(min(avg_online_minutes),2) as min ,
(select round(avg_mins,2) from
(select row_number() over (order by avg_online_minutes asc) as
ranking ,avg_online_minutes as avg_mins from `data`) as temp
where ranking =828934/4+0.5) as Q1,
(select round(avg(avg_mins),2) as median_session_length from(select
avg_mins from
(select row_number() over (order by avg_online_minutes) as ranking
,avg_online_minutes as avg_mins from `data`) as temp
where ranking =828934/2 union
select avg_mins from
(select row_number() over (order by avg_online_minutes) as ranking
,avg_online_minutes as avg_mins from `data`) as temp
where ranking =828934/2+1 ) temp2) as Median,
(select round(avg_mins,2) as avg_mins_Q3 from
(select row_number() over (order by avg_online_minutes desc) as
ranking ,avg_online_minutes as avg_mins from `data`) as temp
where ranking =828934/4-0.5) as Q3,
 round(max(avg_online_minutes),2) as max from data;
```

avg	min	Q1	Median	Q3	max
▶ 11.74	0.00	0.50	1.67	5.00	1605.83

- *Active Rate*
  Active rate is the rate of at least login the game which is means avg_online_minute>0. Here the active rate is 97.9%.

```
/* overall active rate= number of active user/ number of all user */
select round((select count(distinct user_id) from data where
avg_online_minutes>0)/count(distinct user_id),2) as 'log_in(%)' from
data;
```

active(%)
▶ 97.90

- *Very Active Rate*
  Very Active rate is users with avg_online_minutes>5 which is 24.95%.

```
select (select count(distinct user_id) from data where
avg_online_minutes>10)/ count(distinct user_id) as very_active_rate
from data;
```

very_active_rate
▸ 0.2495

- *Percentage of Paying User (PPU)*
  PPU is the Percentage of Paying User (PPU = DPU / DAU), Rate that represents the portion of paying users among all daily active users. Here PPU is 2.41%.

```
/*compute PPU*/
select (select count(user_id) from data where pay_count>0 and
avg_online_minutes>0) /count(user_id) as PPU
from data where avg_online_minutes>0
```

PPU
▸ 0.0241

- *Average Revenue Per Active User (ARPU)*
  Average Revenue Per User (ARPU) is the revenue generated (on average) by each active person using your app. ARPU is $0.69.

```
select sum(pay_price)/count(distinct user_id) as ARPU from data
where avg_online_minutes>0;
```

ARPU
▸ 0.686242

- *Average Revenue Per Paying User (ARPPU)*
  ARPPU is the revenue generated (on average) by each paying user. ARPPU is $28.49.

```
select sum(pay_price)/ count(distinct user_id) as ARPPU from data
where pay_count>0;
```

ARPPU
▸ 28.487393

Answer 1:

By comparing "Brutal Age" performance with Game analytics report, the overall performance of "Brutal Age" is much better than industrial average indicator.

| Key Metrics	Brutal Age	Game Analytics Report
Average session Length	11.74 minutes	7-8minutes for top 25%
Active Rate	97.9%	11%
Percentage of Paying User (PPU)	2.41%	1.8%-3.6% for top
Average Revenue Per Paying User (ARPPU)	28.49	25

Although the paying user rate is at the top role-playing games, the percentage of paying rate is 0.0241 which is low, and it still can be improved.
In the future, the company should investigate the different customer background data and geographic data to offer more user-orientated promotions and strategies.

Question 2: Are These Key Metrics Affected by Different Register Dates?

- *New Users by Register Date*

```
/*count of daily new users */
select count(distinct user_id) as n_DailyNewUser, date(time) as Date
from data group by date(time) order by date(time);
```

Number of daily new user

Series "Number of daily new user" Point "2018-03-10"
Value: 114035
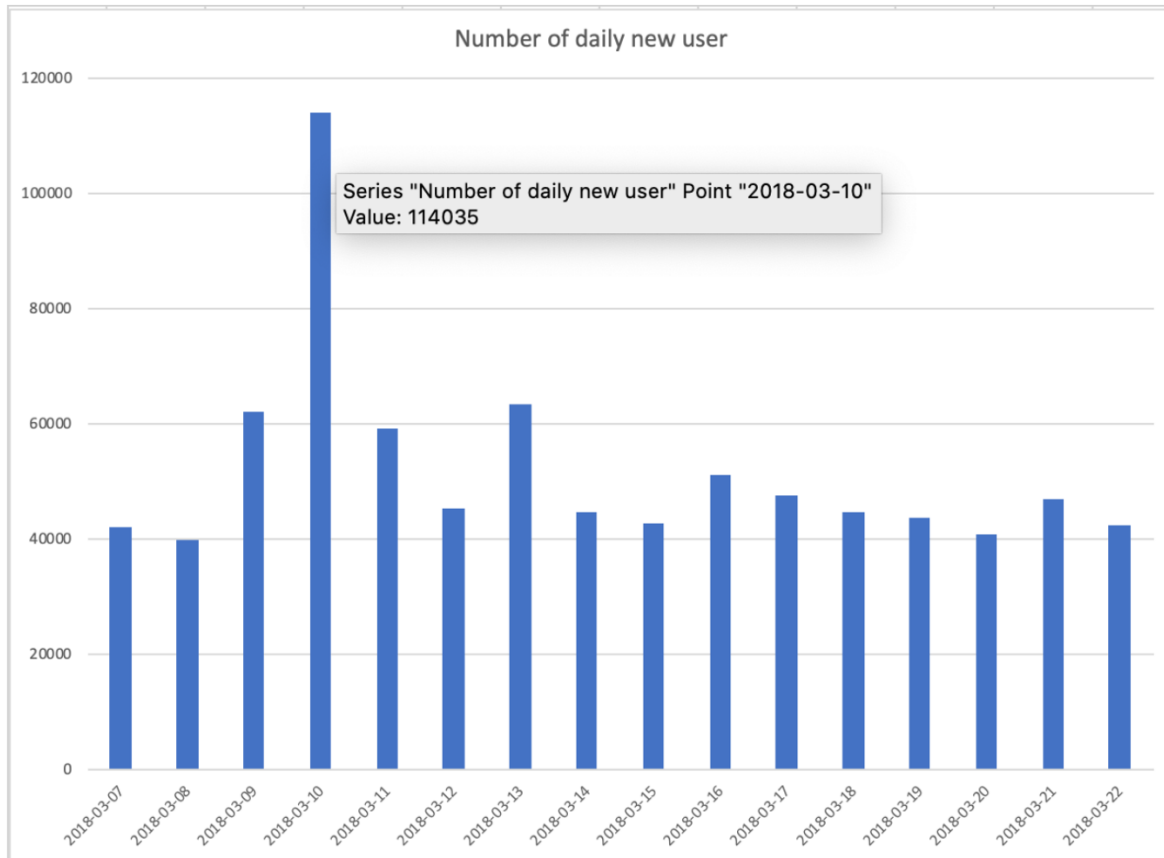
From above histogram, we can clearly tell that on 2018-03–10, the number of daily new users almost doubled as compared with the rest of 15 days. This sudden increase is abnormal. We will investigate what makes this one-day new user uprush. we can also figure out after 2018–03–10, the number of daily new users gone back to a normal level. We should investigate what makes this data back to normal so quickly. There are another two little peaks on the new register user number on 2018–03–13 and 2018–03–16.

Let's figure out why there were a sudden increase on 03-10 and why there are other two little peaks on 03-13 and 03-16.

From marketing aspect: In MySQL, DAYOFWEEK function result, 1 represents Sunday, 2 represents Monday. From SQL result, 2018–03–09 is Friday,03–10 is Saturday,3–11 is Sunday, ,3–13 is Tuesday, 3/16 is Friday. 3/9,3/10,3/11, 3/16 is the weekend. The result supports the hypothesis; However, 3/13 is not a weekend or a holiday.

```
select '2018-03-09' as Date, DAYOFWEEK('2018-03-09')union
select '2018-03-10', DAYOFWEEK('2018-03-10') union
select '2018-03-11', DAYOFWEEK('2018-03-11') union
select '2018-03-13' ,DAYOFWEEK('2018-03-13') union
select '2018-03-16' ,DAYOFWEEK('2018-03-16');
```

Date	DAYOFWEEK('2018-03
2018-03-09	6
2018-03-10	7
2018-03-11	1
2018-03-13	3
2018-03-16	6

From product aspect: is the game has a special activity on these days?

By searching Brutal Age's Facebook post, it is found that on 03/09/2018, there is a new Brutal Age released version. The new version included a new function:3D display Berserker which is the key character in Brutal Age. The new 3D display model attracts lots of new users. The evidence supports the hypothesis.

It is found that crazy mammoths will appear near your Outpost Portal on 03-12. This increase is going on.

On 3/16, Brutal Age will hold St. Patrick's Day special events from March 16 to March 19. During this period, exclusive items, decorations, and monsters will be available! There's enough evidence to support the hypothesis.
The increase in new users on 3/9,3/10,3/11 is because there's a new 3D display model and these days are weekends. The increase in new users on 3/13,3/16 is caused by a special event as well.

In addition, daily new users may be higher at weekends than on weekdays. However, 16 days of data is not enough to assert the result. We need more data to do weekend vs weekday new user analysis.

- *New Users During Different Time Period for Each Day*

```
/* Hourly new user from 3/7 to 3/22*/
select count(user_id) as n,DATE_FORMAT(time,'%m-%d') as
day,DATE_FORMAT(time,'%H') as hour
from `data` group by DATE_FORMAT(time,'%m-%d')
,DATE_FORMAT(time,'%H') order by day, hour;
```
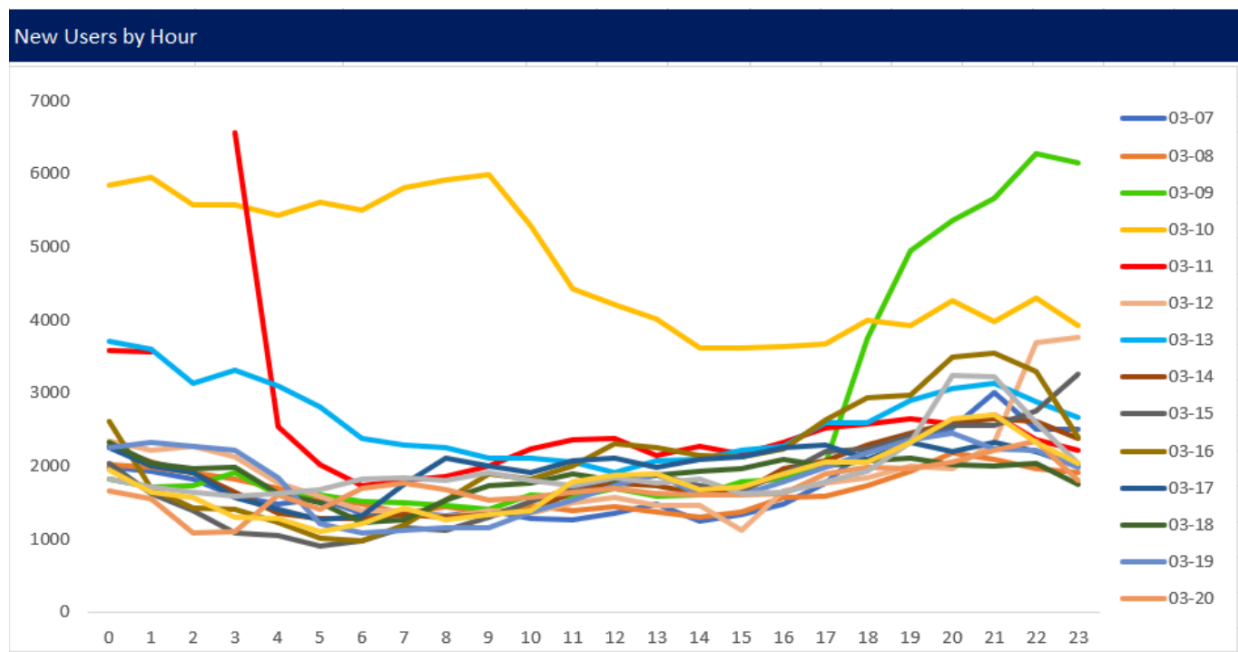

From above line chart, during most of these days, the number of new users is increasing from 17:00–22:00. The peak number of new users usually at 21:00. Based on the number of new users, we can know most users play this game at 21:00. Fruthermore, lines of 3/9, 3/10, 3/11, 3/13 are different from other lines. The number of new users is rapidly increasing from 3/9 17:00 to 3/11 6:00. After 3/11 6:00, the number of new users slows down and becomes normal as usual.

After investigation, On March 12 at 14:00 UTC there were crazy mammoths will appear near Outpost Portal. Therefore, Let's look back to Figure 5.2 to see if the increase in new users on 3/13 happens from 3/12. And the line for 3/12 is increasing from 21:00 until 23:59. This increase is going on.

- *Average Session Length by Registration Date*

```
/*average session length of new users group by date */
select round(avg(avg_online_minutes),2) as avg_session_length,
date(time) from `data`
GROUP BY date(time) order by date(time);
```


Average Session Length for New Users by Register Date
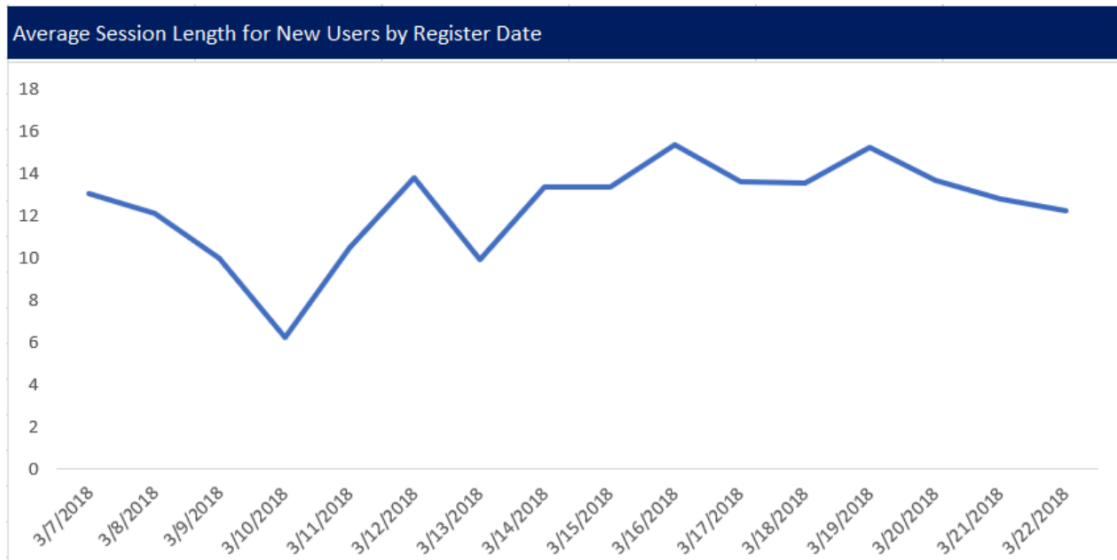
From above line chart, Average session length for the users who registered on 3/9, 3/10, 3/11, 3/13 is less than that for those users who registered on a usual day. The average session length for the users who registered on 3/16, and 3/19 is higher than other days.

we can know the 3D display model can attract a large number of users but it cannot bring more active users. The game should think about how to keep more users attracted by the 3D display.

In addition, the event which provides exclusive items, decorations, and monsters can improve the stickiness of this game.

- *Active Rate by Register Date*

```
/* active rate by register date  */
select a.date,  round(a.n/b.n*100,2) as log_in from
(select count(distinct user_id) as n ,date(time) as date from data
where avg_online_minutes>0 group by date(time) order by date(time))
as a left join
(select count(distinct user_id) as n, date(time) as date from data
group by date(time))as b on a.date=b.date;
```

- *Very Active Rate* (the rate of User with average online minutes > 5)

```
select a.date, round(a.n/b.n*100,2) as 'very_active_rate%' from
(select count(distinct user_id) as n ,date(time) as date from data
where avg_online_minutes>5 group by date(time) order by date(time))
as a left join
(select count(distinct user_id) as n, date(time) as date from data
group by date(time))as b on a.date=b.date
```

Active Rate by Register Date
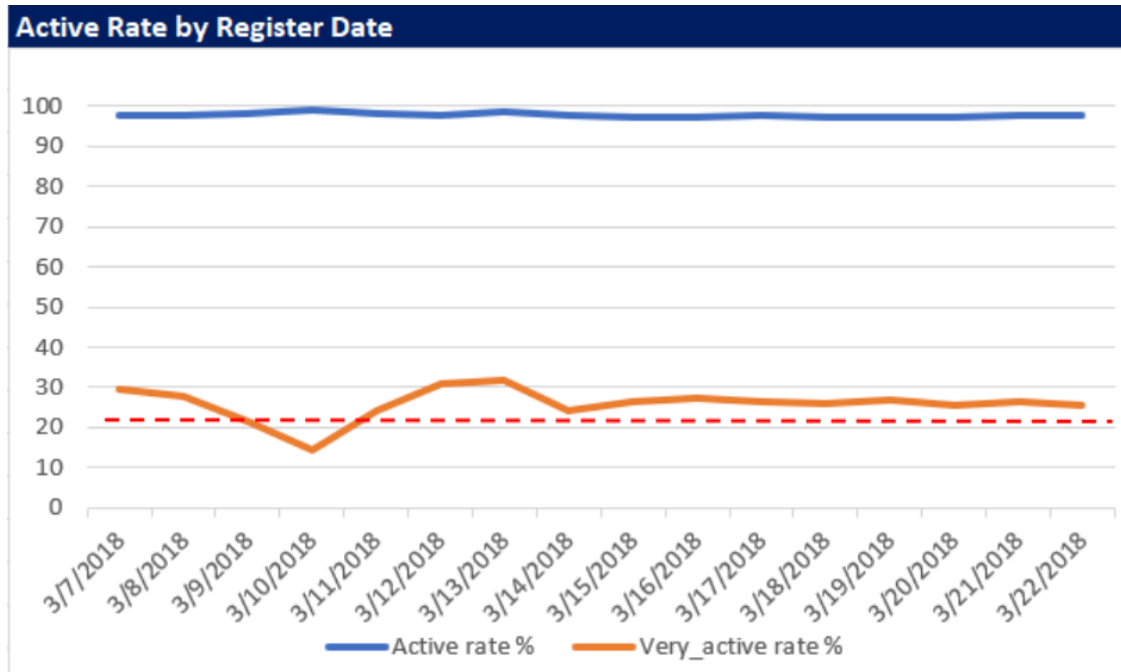
From above plot, we can tell the active rate is almost the same for users registered on different dates. When it comes to the very active rate (avg online minutes >5), 3/10's is lower than other days. The very active rate for 3/12 and 3/13 is higher than that for other days. we can know different events can bring different effects. New appearance, new 3d model can attract more users to register the game. The wild hunting events can bring more active users and help to improve the stickiness of the game.

- *Percentage of Paying User (PPU) by Register Date*

```
select a.Date, a.n/b.n as PUR from
(select date(time) as Date, count(distinct user_id) as n from data
where pay_count>0 group by date(time) ) as a
left join (select date(time) as Date ,count(distinct user_id) as n
from data where avg_online_minutes >0 group by date(time)) as b on
a.Date=b.Date
```

PPU By Register Date
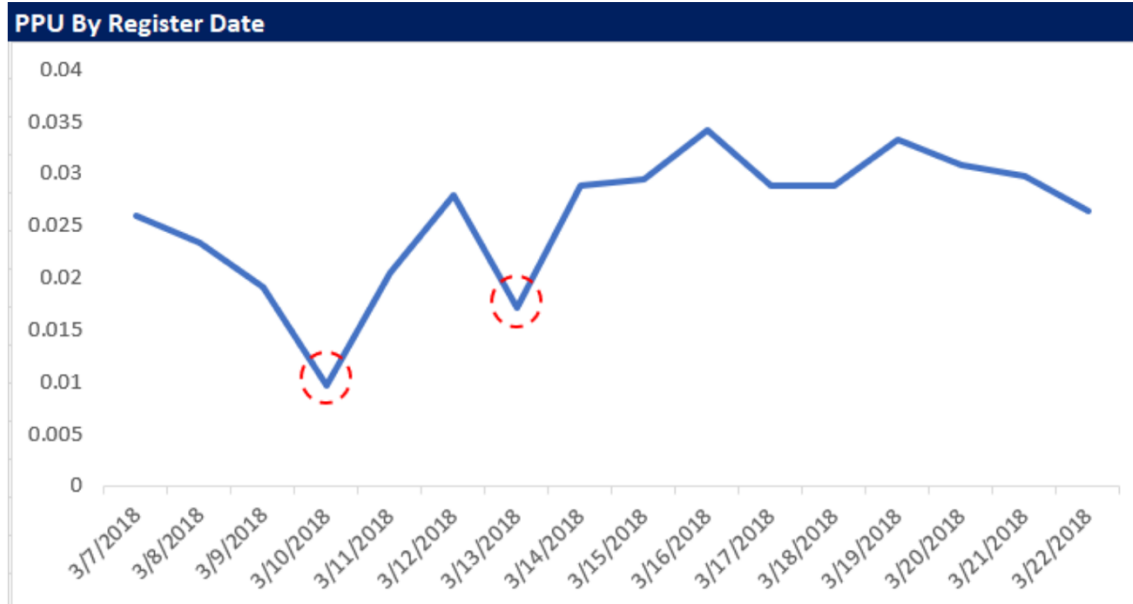
From above plot, PPU for 3/10 and 3/13 is relatively low when compared with other days. However, PPU for 3/16 is a little bit higher than other days, which means exclusive items, decorations, and monsters available on that day can stimulate user purchase. The new 3D model and wild hunting event won't bring more paying users in percentage. The release of exclusive items can arouse the new user's purchasing desire.

- *Average Revenue Per Active User (ARPU)*

```
select date(time) as Date ,sum(pay_price)/count(distinct user_id) as
ARPU from data where avg_online_minutes>0 group by date(time);
```


ARPU By Register Date
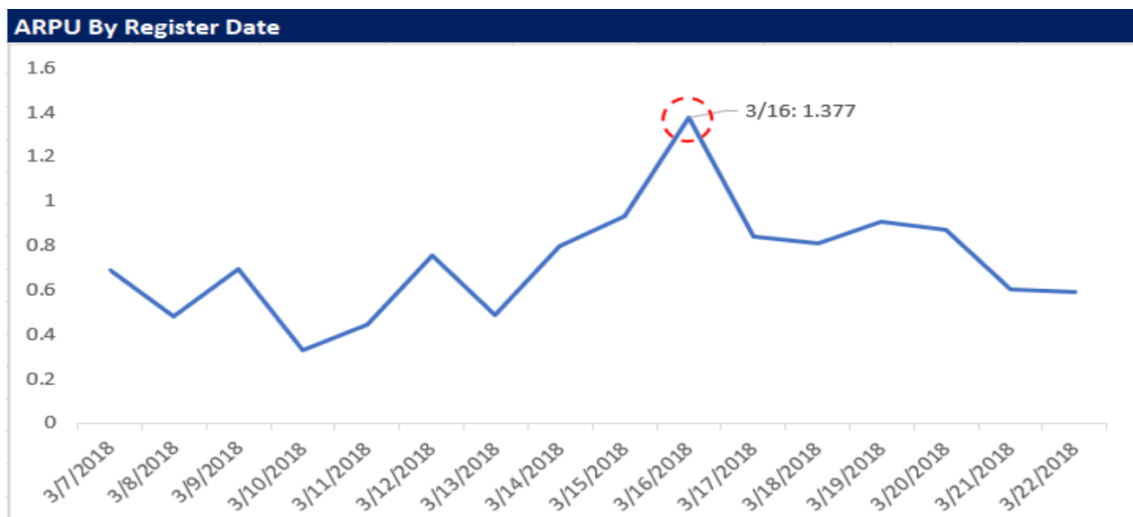
As above plot shown, ARPU is extremely high for 3/16, which implies the event on 3/16 is very successful. Users love these exclusive items, decorations, and monsters available on that day. They would like to spend a lot on them.

- *Average Revenue Per Paying User (ARPPU)*

```
select date(time) as Date, sum(pay_price)/ count(distinct user_id)
as ARPPU from data where pay_count>0 group by date(time);
```

ARPPU By Register Date
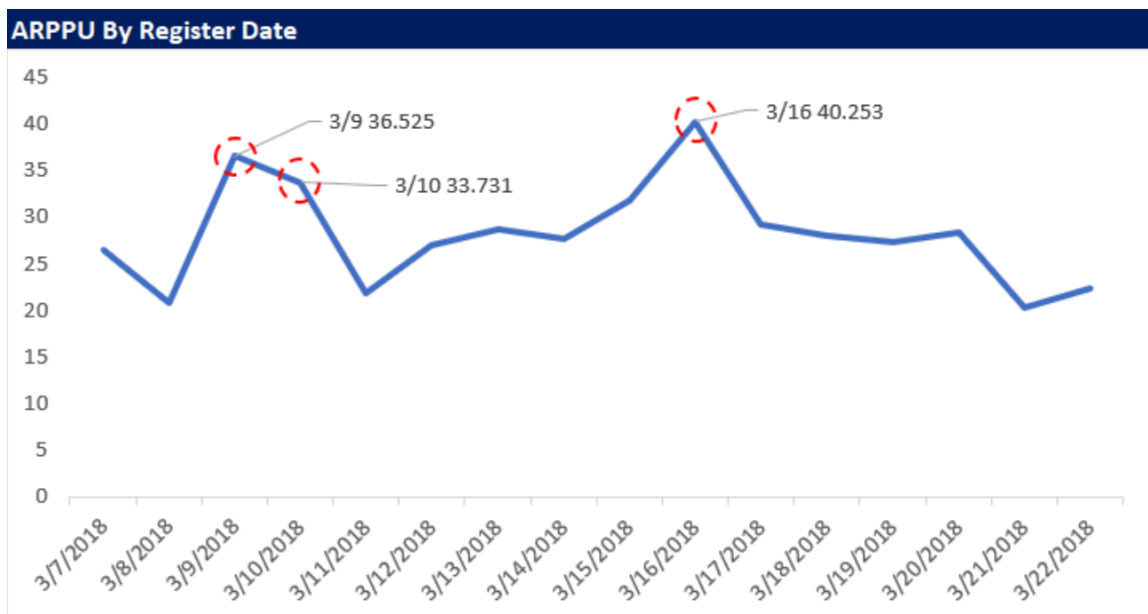
The ARPPU for 3/9 and 3/10 is as high as 3/16. It implies that there are more high-quality users who prefer to spend money on the game register on 3/9 and 3/10.

VI Conclusion

The new appearance or 3D display model can attract a great number of users; Although it cannot last for a long time, and less active users as well as paying users among those users attracted by these new appearances, there are more high-quality paying users.

The wild hunting event cannot bring more paying users and high quality paying users. However, this event can improve the stickiness of the game. Those users registered the game for this event have a higher active rate.

Exclusive items, decorations, and monsters available on 3/16 can not only attract new users but also can arouse user's purchasing desire.

When the game releases a new appearance or 3D model, it can also provide exclusive items, decorations. The high-quality users attracted by new appearance has a great purchasing power. They would like to purchase more than other users.