

Supercell Report

Hung Nguyen

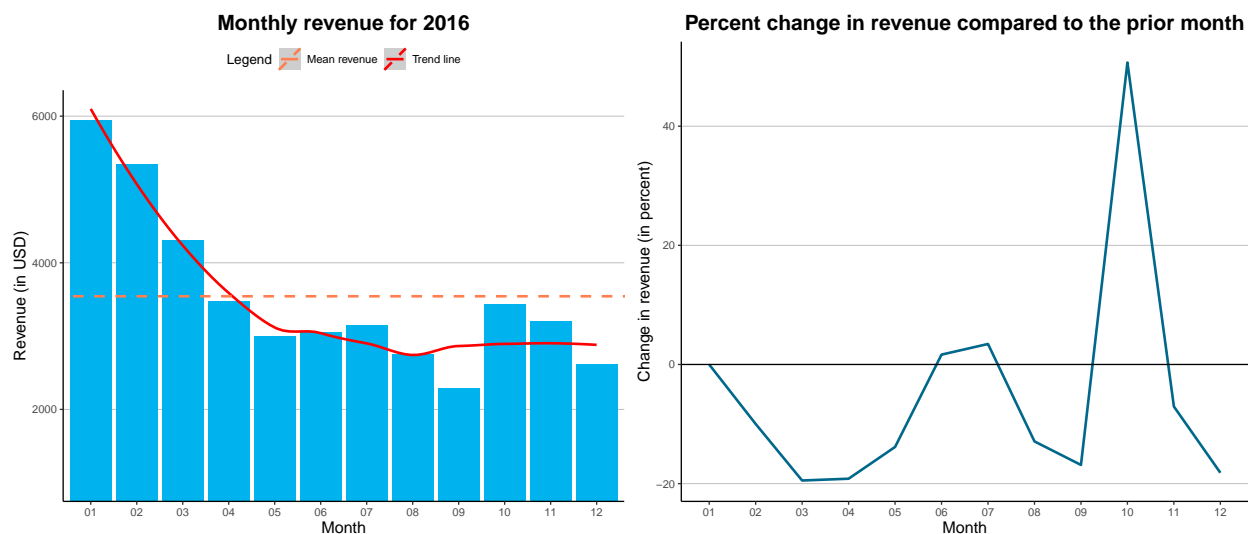
2022-12-30

Warmup: Revenue analysis

Revenue trend analysis

We can examine the time series revenue and the revenue change during 2016 to analyze its trend:

- **Monthly revenue for 2016** measures the total revenue from users' purchases for each month.
- **Percent change in revenue compared to the prior month** measures the change in revenue (in percent) compared to the previous one for each month. (For January, it is assumed that the percent change is 0)



From the first graph, we can first see that the most noticeable peak revenue is in January, which is at climax with around USD 6000 in revenue. Meanwhile, the revenue trough occurs in September with less than USD 3000 generated. We can also observe that the only times the revenue remains above the mean are the first three months of the year: January (around USD 6000), February (around USD 5500) and March (around USD 4300). It is also worthwhile to notice the change from month 9 to 10, where we can clearly observe a great bounce back from the revenue trough by over USD 1000 gained more in October, compared the prior month. Apart from that, during May to December, the revenue in general witnesses no other significant change. It gradually falls from January until reaching below the mean, then remains quite stable since then with only small business cycle fluctuations. This trend is also suggested by the trend line, where the revenue begins from January with the highest peak and has some falling actions until becoming relatively stable below the mean revenue since May.

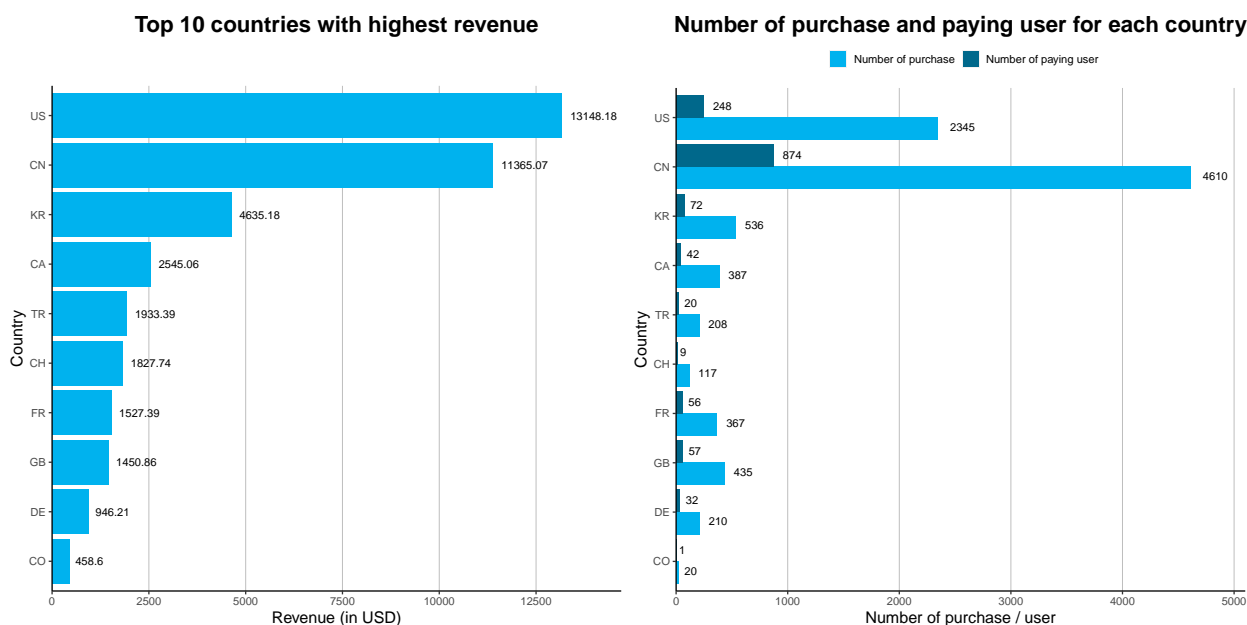
From the second plot, we can notice that the changes are stable for the most part as it would take some time remaining negative (or positive) before changing to the opposite side again. This is not to mention the rise from month 9 and 10, where it is observed as the peak percentage change with around 50% of change; this is expected as in the previous plot a bounce back from a revenue trough during this period is already mentioned. In addition, we can also see that the change in revenue spends a longer duration being negative than positive, indicating that the revenue throughout the year is more likely to fall than to rise.

Geographic split of revenue:

In this section, we are going to assess the geographic split in revenues with respect to the top 10 countries with the highest revenue and with respect to continental and product type aspects. Analyzing the top 10 most revenue generated countries will help identify specifically the most potential markets to be taken care of. In order to look at statistics in a greater scope because of the high volume of countries, we may look into the most potential continents along with the revenue from each package in these market.

We can first analyze the revenue from the top 10 highest revenue generated countries as follows:

- **Top 10 countries with highest revenue** measures the revenue for each of the top 10 highest-revenue countries.
- **Number of purchase and paying user for each country** has two quantities of interest: *Number of purchase* measures the number of purchase made by users (can be multiple times), and *Number of paying user* measures the number of different user within that country purchasing at least one package. The selected countries are the same as the top 10 countries with highest revenue.

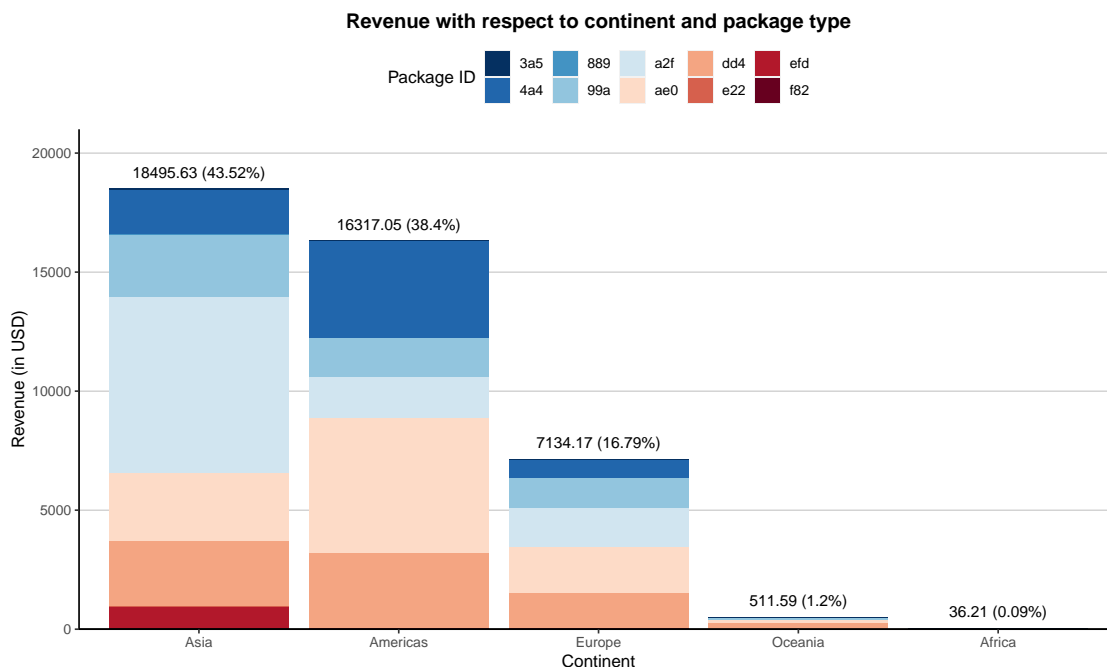


It can be seen from the second plot that the two countries with the most significant difference among the others having the highest number of paying user, purchase, as well as re-purchase are US (top 2) and CN (top 1). In the first plot, as expected, it is clear that the top two countries with the highest revenue are also US and CN, though the revenues generated by CN are relatively smaller compared to that of US. This shows that US is a high monetization market, while CN is a high volume region. Nonetheless, the plot strongly suggests the exceptionally higher demand for the products within both countries.

Although the sales are not as impressive as CN and US, country KR, having slightly the same paying users and purchases as other countries excluding the top 2, generates a noticeably higher revenue than the lower

ones and place itself in the top 3 highest revenue generated country. It is also worthwhile to notice that country CH and CO, having a significantly smaller portion of paying users and purchases than others, can generate such a revenue to place themselves in the top 6 and 10, respectively. This might be explained that the packages with higher price are more popular among KR, CH, and CO and the average revenue per paying user (ARPPU) is greater.

We can also look into the the revenue with respect to each continental and each package to assess the data from a greater scope. For the continental split, a mapping of each country to the corresponding continents are used (the info can be found in the References section). For the packages, we only take the first 3 characters of the package ID `package_id_hash`, because it is enough to differentiate between the packages using that information; moreover, this can improves the visualization by avoiding displaying too much characters.



With over double the revenue generated compared to the top 3 continent Europe, Asia (top 1) and Americas (top 2) generate the most distinctive and highest revenue, with 43.52% and 38.4% contribution to total revenue. This is expected, as CN and US, being the top 2 highest contributors in terms of revenue and number of paying user, are located in Asia and Americas, respectively. Meanwhile, it is clear that Oceania and Africa only contribute very little (only 1.2% and 0.09%), showing that the market at these continents are quite small.

In addition, we can also observe the popularity of each package within each continent. For Asia, `a2f` brings about the most revenue among other packages, with around USD 7000 (more than a third of total revenue in Asia) among other packages, which is more than the revenue of the same product in any other continent. The same thing goes to Americas, where the sales of the package `ae0` and `4a4` are the highest ones and the consumption of theses products are higher than any other continent. Apart from that, it is also worthwhile to notice the revenue from the package `99a`, `dd4` are relatively the same within the top 3, while the remaining packages `3a5`, `efd`, `f82`, `e22` and `889` only contribute considerably smaller portions of revenue, or do not contribute at all.

Task 1. Lifetime Reveue (LTV)

LTV definition

Lifetime revenue is a metric that captures the total amount of revenue a user spends on company's business during the entire course of the user's relationship with the company.

Predict players LTV during their first week in-game:

The "top-down" monetization approach is used to calculate LTV, where it aims to measure the revenue that customer spend over their whole lifespan in the game. The following three assumptions were made for this problem:

- The main priority is revenue. Therefore, the only segment of user taken into account is paying user. In addition, the metric Average revenue per paying user (ARPPU) is used instead of the commonly used metric Average revenue per user (ARPU) for this method. As the proportion of paying user only counts very little compared to non-paying ones according to the data (Only 1.37%), it is better to use ARPPU because it disregards the large noise from users that does not generate any revenue for the company and more accurately models the actual population that contributes to the revenue.
- Most of the LTV of a user is generated within early in-game time of the user and then decays over time. Thus, Logarithmic Regression will be used to model this behavior.
- LTV will be modeled over a period of 180 days. As there is little information on the company status and the genre of the game itself to choose a suitable period, I assume the game is casual and the 180-day period is chosen because it is common for most casual games, according to Seufert E. (2014).

From the data, ARPPU is extracted by daily cohort to form a table. The first column **reg_date** contains registration date, the second column **registrations** contains the total number of user creating their account for the first time during that date, while the columns **1,2,..., 361** contain users' ARPPU from 1, 2,..., 361 days from their corresponding registration date. A snapshot of daily ARPPU is provided as follows:

reg_date	registrations	1	2	3	4	5
2016-01-01	18	1.650000	2.940000	1.470000	4.306667	NA
2016-01-02	35	4.481667	2.456667	6.638000	4.150000	9.230000
2016-01-03	23	10.956667	19.226667	73.980000	0.360000	13.014000
2016-01-04	20	15.458571	6.640000	NA	8.006667	12.320000
2016-01-05	17	8.290000	1.840000	NA	2.302500	2.456667
2016-01-06	19	1.190000	9.230000	5.523333	9.843333	6.153333

From this, we calculate the weighted average ARPPU of the user cohort for each day 1, 2,..., 361 days, with the values are the ARPPU from the rows of each column **1,2,..., 361** and the corresponding weights are the **registrations** amount. A snapshot of 5 first days corresponding to the table above can be given as:

1	2	3	4	5
5.863844	5.177217	6.511579	4.984194	4.913086

As mentioned above, we will use Logarithmic Regression to model the behavior of daily ARPPU. The equation of the function is given as follows:

$$y = a \ln(x) + b,$$

where x is the number of days since installed, and y is the projected LTV

From fitting a Logarithmic Regression model into the 361-day ARPPU data, we obtain the following coefficients:

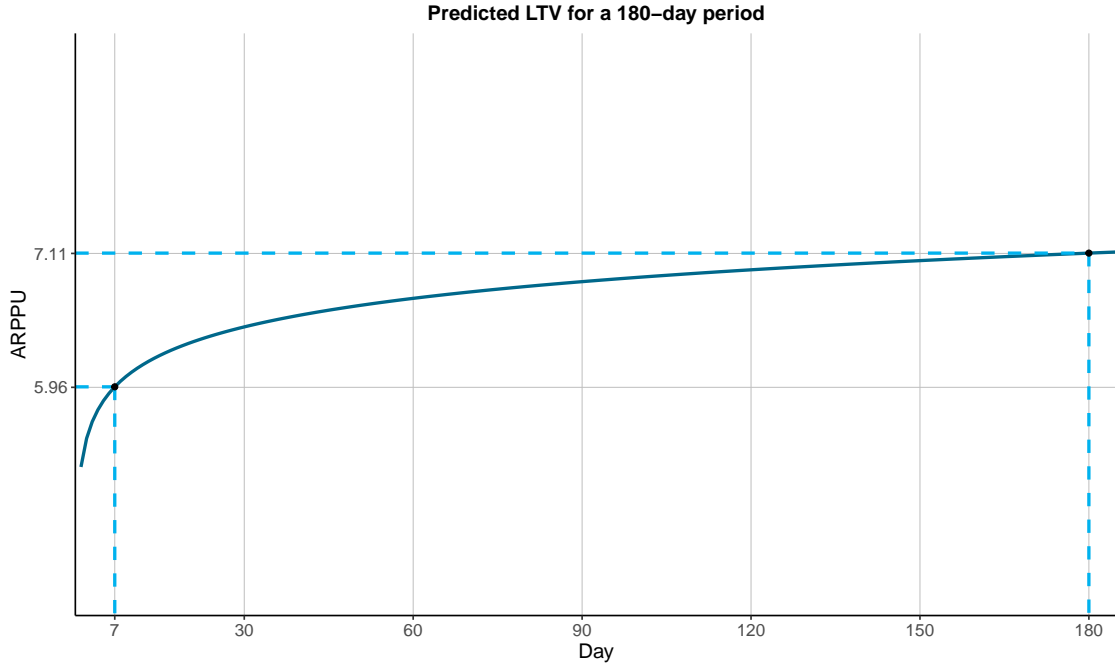
intercept	slope
5.275918	0.3537022

Therefore, we obtain the following curve of best fit:

$$y = 0.3537022 \ln(x) + 5.275918,$$

where x is the number of days since installed, and y is the projected LTV

Using this equation, we can model the approximate spending behavior of a user and predict the cumulative ARPPU as well as the LTV for the next 180 days since that user's first registration for the game account:



The line graph models the cumulative ARPPU, or the average spending behavior of a (paying) user throughout his estimated lifetime of 180 days. At day 180, user retention period ends and the LTV is 7.11. Therefore, during the first week (7 days) since their first install of the game, the proportion of total LTV (for an 180-day period) a user generates is:

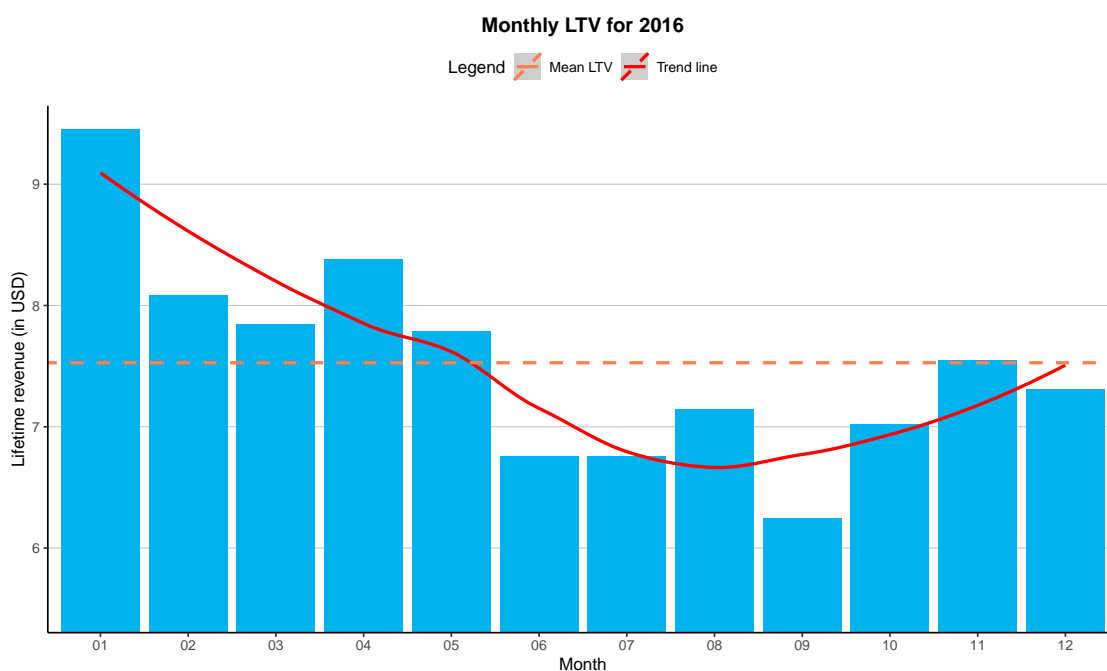
$$\frac{5.96}{7.11} = 83.83\%$$

Analyze monthly LTV from January to December

In this section, in order to have a general assessment for the LTV during 2016, we will look into the monthly LTV from January to December. In addition, the monthly LTV statistic will be analyzed separately for different platforms (iOS and Android) through user segmentation. Users are acquired through different platforms and their spending behavior can be distinctive, and this suggests the need to analyze the trend for LTV separately through different channels.

The “top-down” monetization approach is again used to calculate LTV of an average paying user for each of the 12 months. Under the same assumptions, this method will still apply Logarithmic Regression to predict LTV given a user retention period of 180 day. However, as LTV are being calculated for each month, we will extract the data from the daily cohort ARPPU and calculate the weighted average ARPPU for each month instead of for all months as in the previous problem. With each month from January to December, a Logarithmic Regression model will be fitted based on the daily cohort ARPPU of previous months (including the current one). With each fit, we obtain the regression coefficients to compute the corresponding 180-day LTV.

The following column chart of monthly LTV is obtained:



While the LTV hits the bottom in September with around USD 1.5 below the mean, it peaks higher at around USD 2 above the mean. It is also worthwhile to notice that the most noticeable LTV changes during this year are mostly downfalls, occurring during from January to February, May to June, August to September with around over USD 1 lost in customer value. For the first half (January to May) of the year, though the LTV has some considerable falling actions, it remains above the mean the whole time. In the opposite, during the rest of the year (June to December), the LTV is below the mean line, with some of the months in the middle of the year like June, July and September being the (negatively) furthest from the mean LTV. This is with the exception of November, which is the only month during the second half having the LTV value larger than the mean, though very slightly. From the trend line, we can also observe the overall trend of LTV throughout January to December is downward. However, if we analyze the changes of LTV through the column plot, the trend of LTV is actually relatively similar to the monthly revenue, remaining at the peak at the beginning of the year, gradually decreasing until May, then remaining stable until suddenly reaching a low bottom in September and bounce back in October. This shows that the LTV fluctuations

are somewhat sensitive to the revenue fluctuations and the changes in LTV therefore can be addressed to changes in total in-game revenue.

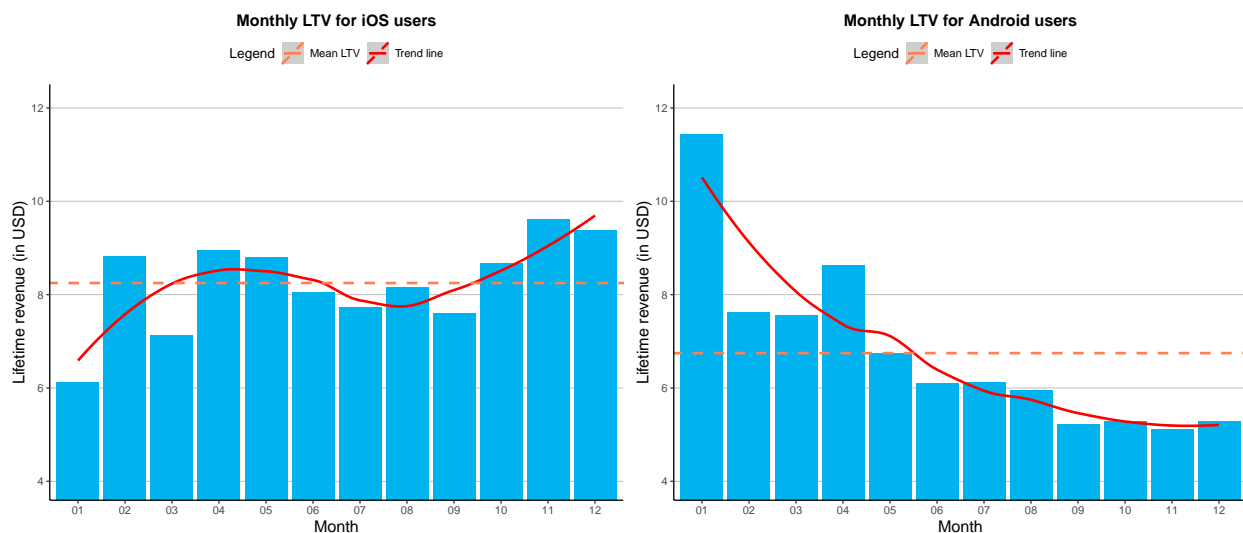
We may also look at the LTV with user segmentation by platforms applied to understand the LTV of users acquired through different channels (iOS and Android).

First, we have to segment the user by platforms. The daily cohort ARPPU table above is separated by `created_platform` and a snapshot can be provided as follows:

reg_date	registrations	created_platform	1	2	3	4	5
2016-01-01	14	Android	1.650000	2.940000	1.346667	4.306667	NA
2016-01-01	4	iOS	NA	NA	1.840000	NA	NA
2016-01-02	24	Android	4.721818	2.456667	5.527500	2.760000	3.68000
2016-01-02	11	iOS	1.840000	NA	11.080000	5.540000	14.78000
2016-01-03	14	Android	4.422500	1.100000	73.980000	0.360000	14.91333

After performing the same steps for each platform Android and iOS separately to generate monthly LTV calculations from January to December, the following plots and table are obtained:

- **Monthly LTV for iOS users** tracks the monthly LTV of a paying user using iOS devices from January to December.
- **Monthly LTV for Android users** tracks the monthly LTV of a paying user using Android devices from January to December.
- **Number of user across platforms** measures the number of paying user for each platform.



Number of user across platforms

created_platform	user_count
Android	1151
iOS	398

With only around 26% of total players, iOS users in general have greater LTV than Android users (74% of total players) throughout the year. This is visible from the mean lines from both graphs, where it is clear that the mean LTV of iOS users are around USD 1.5 higher compared to the mean LTV of Android player.

It is also worthwhile to notice the trend of LTV for both platform. The monthly LTV of Android players slightly resembles the overall LTV by gradually falling since January and remaining stably low throughout the year. Meanwhile, the monthly LTV of iOS player has the almost opposite trend, where it is observed to rise considerably from January and remain at high level of LTV. Although the peak LTV of iOS user are not as high as that of Android user, the trough LTV of iOS user is not as low; moreover, for most of the time, the LTV of iOS user remain higher. This shows that even though iOS users only have a smaller portion of total player population compared to Android users, they are more monetarily valuable and are more likely to spend.

References

- Olteanu, A., *Country Mapping - ISO, Continent, Region*. <https://www.kaggle.com/datasets/andradaolteanu/country-mapping-iso-continent-region>
- Seufert, E., *Leveraging Customer Lifetime Value for user acquisition campaigns*. 2014. <https://www.slideshare.net/AppLift/ltv-webinar-presentation>