Epic Games Assessment

Role: Marketing Performance Manager

Applicant: Casey Moroney

10/24/2024

Exercise 1

Prompt:

“On a simple webpage that allows the download of a game a test is running to determine which variant (A or B) should be rolled out. Please make a recommendation based on the data attached.”

Ex. 1: A/B Test

Assumptions:

- Page “A” is the business-as-usual page and “B” is the experiment / variant
- Users are randomly assigned to either page A or B upon visit
- There is meant to be a 50/50 split between the two pages

Results:

- Variant B had a -5.17% lift in conversion rate*
- This difference is **not significant**, meaning there is statistically no difference between the two landing pages

Recommendation:

- Review experiment design
  - If this experiment was meant to be a 50/50 split, something may have gone wrong as a sample ratio mismatch exists.
  - If the test was designed so that only 30% of total traffic is directed towards the the variant, the test may need to run longer in order to see a significant result.
- If re-running the test for a true 50/50 split or for a longer period of time with a 70/30 split is not feasible, I would recommend keeping variant A the “business-as-usual” page as we do not have evidence that variant B would perform better.

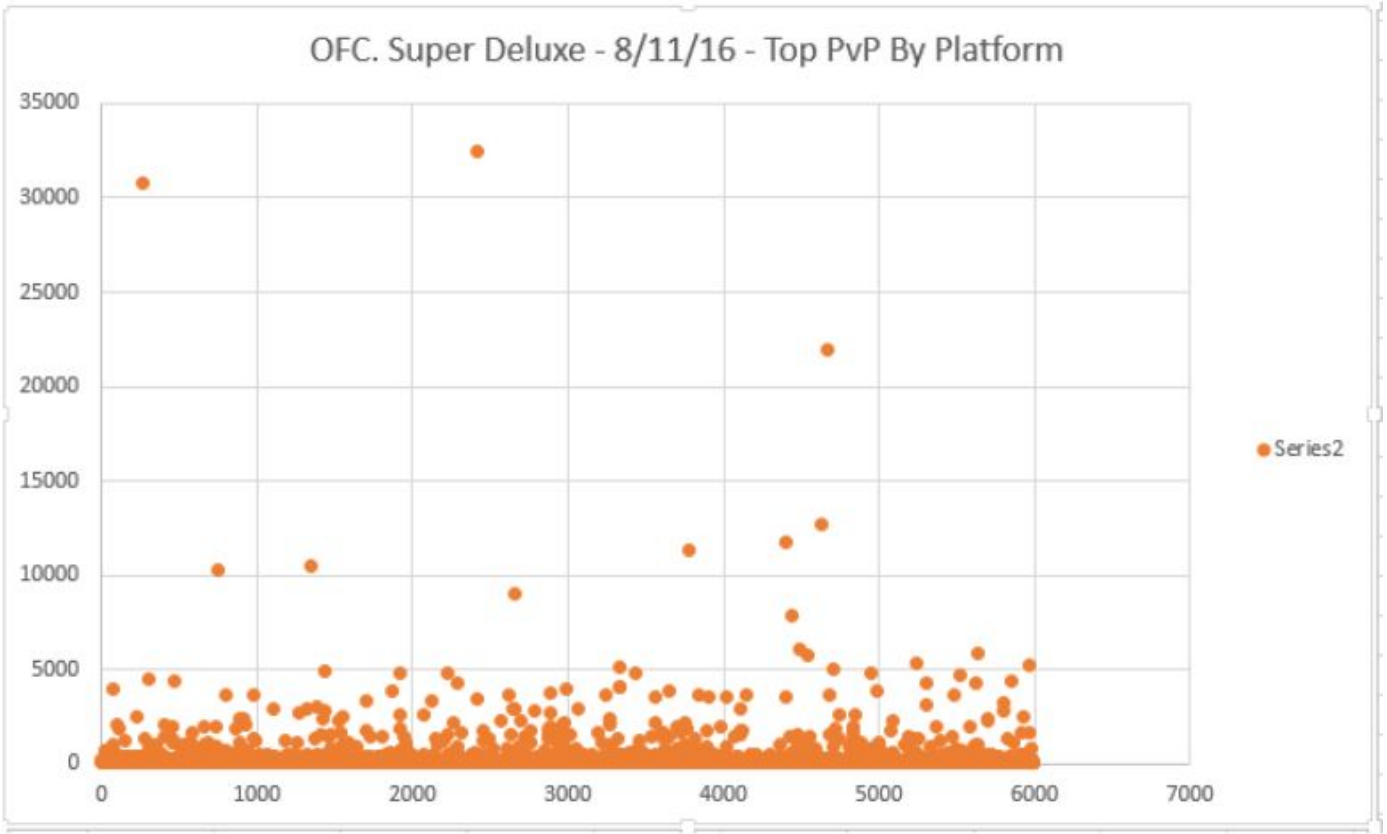
Variant	Visits	Downloads	Conversion Rate
A	271,060	2,107	0.777%
B	117,485	866	0.737%

* Conversion rate = total downloads / total visits

Exercise 2

Prompt:

"What's wrong with these two visualizations and how would you fix them or revisualize them?"



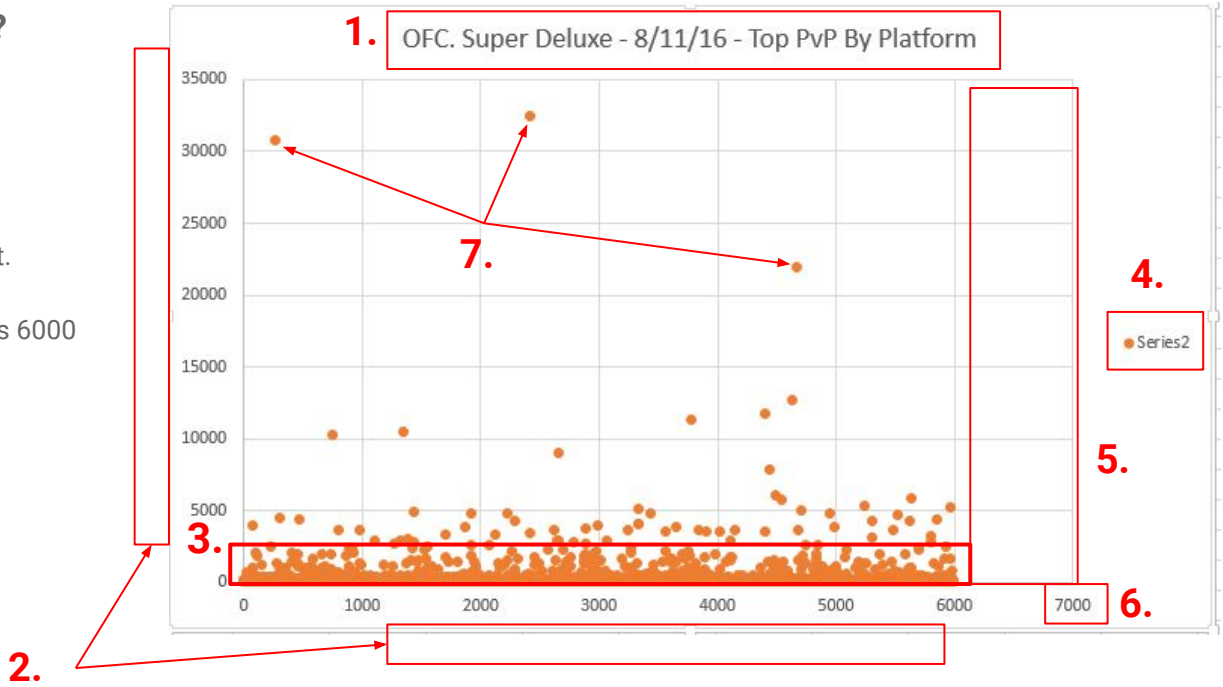
Ex. 2, Visualization 1

Ex. 2: Visualization 1

What's wrong with this visualization?

TL;DR - Everything.

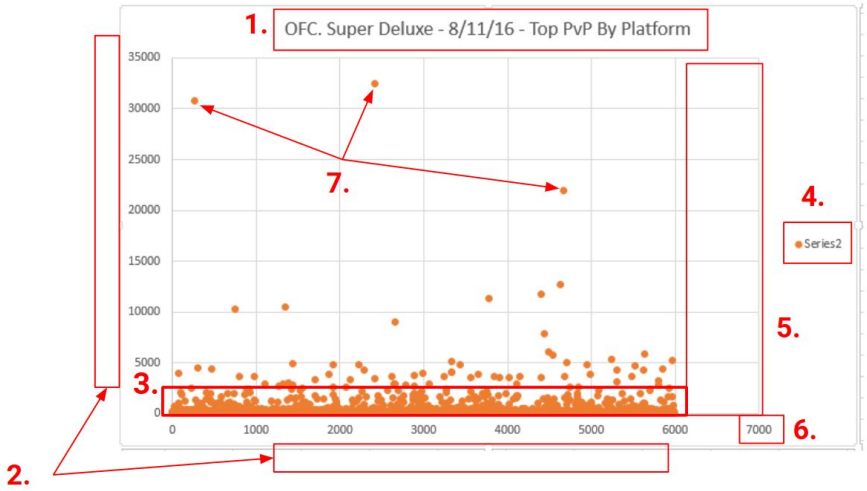
1. The title is unclear.
2. Missing x- and y-axis labels.
3. Data is crowded and difficult to interpret.
4. The legend "series2" is uninformative.
5. There are no data points between values 6000 and 7000 on the x-axis, taking up space without adding value.
6. Numbers are not formatted.
7. Extreme outliers are not highlighted or explained.

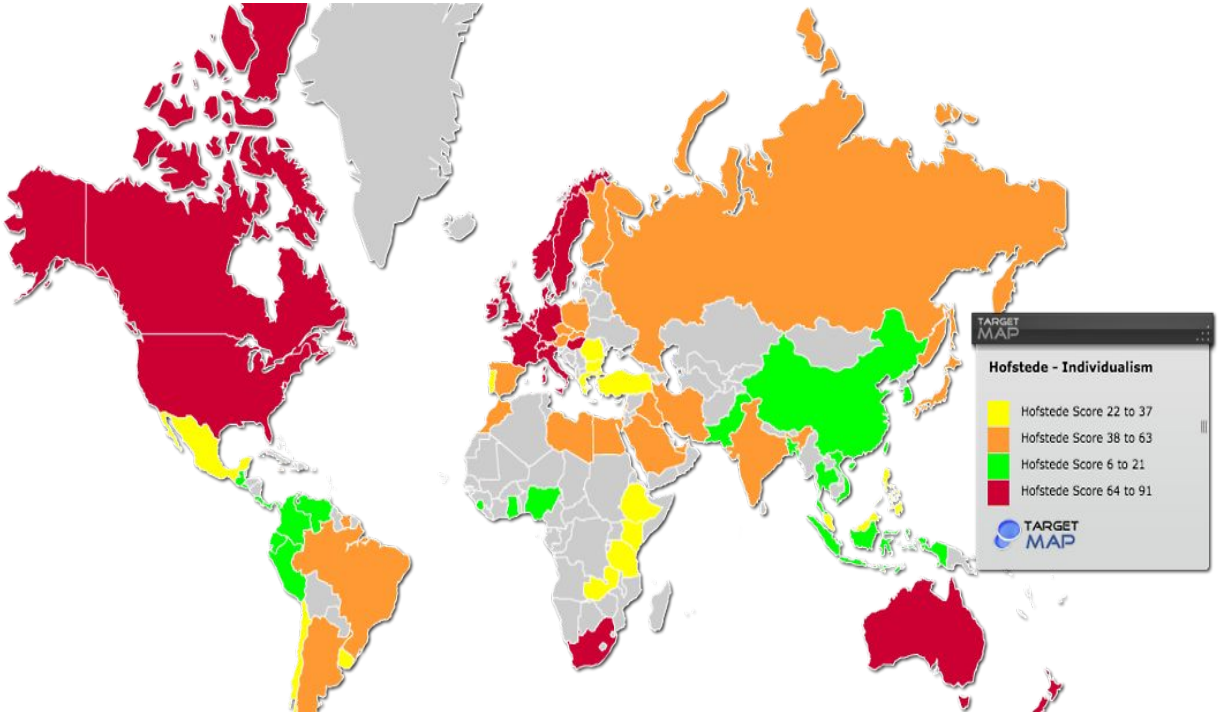


Ex. 2: Visualization 1 (cont.)

How would I fix or revisualize them?

1. A more concise title would make it easier for the viewer to understand. If the abbreviations (OFC., PvP) are kept in the title, they should be defined elsewhere on the page. I would also be more concise - Top PvP what? Top PvP *game*?
2. Add axis titles so the viewer knows what they are looking at, otherwise the visualization is essentially useless. What metrics do the axes represent? Points, dollars, purchases, users, etc.
3. The data is crowded between y-values 0 and ~1000. This suggest there may be many y-values equal to 0, which may be important to note and potentially remove from the graph, but if not a different visualization could be more appropriate. Using binning, a box-plot, or density plot could make the data more readable and trends easier to spot.
4. The title suggests the data is grouped by platform, but only one color exists in the legend. I would group the data by platform (assuming that data is available), then color the data points accordingly and adjust the legend to define the groups. If there is only one group, then there is no need for a legend.
5. Reducing the scale of the x-axis to 6000 (or just over, as to not cut off the data point visuals) would make better use of the page and allow the data to be seen more clearly.
6. I would format axis numbers with commas or currency to make them easier to read.
7. Add labels for extreme outliers to give more context and preempt the inevitable question from the viewer.



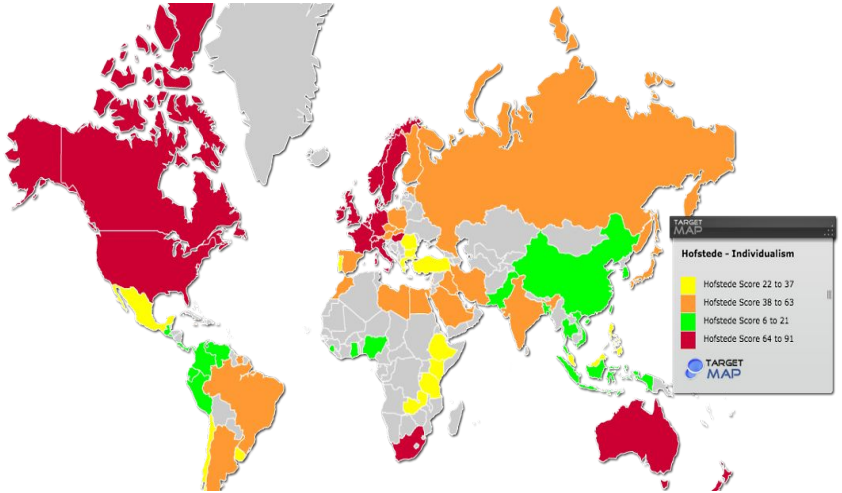


Ex. 2, Visualization 2

Ex. 2: Visualization 2

What's wrong with this visualization? How would I fix or revizualize them?

1. The chart has no title.
    a. Self-explanatory - adding a title lets the viewer know what they are looking at.
2. The colors are arbitrary. They do not intuitively describe whether the country has a high or low Hofsteded-Individualism score.
    a. This could be fixed by using a monochrome gradient with low scores starting light and high scores being dark. This way the viewer more intuitively understands the different shades in relation to the Hofstede - Individualism score.
3. Countries are not labeled.
    a. Viewers that are not familiar with a countries shape would benefit from seeing the country names as labels.
4. The legend is out of order.
    a. It would be more intuitive to order the legend from low to high, starting with "...6 to 21" at the top.
5. No definition for what the data represents and no source noted.
    a. I would add a definition of what the metric represents and link to a source to where the data came from at the bottom of the chart.
6. No explanation for greyed out areas.
    a. If there is missing data or there is a reason for not having data in the greyed out countries, it should be explained somewhere on the chart.



Exercise 3

Prompt:

“You are a Data Scientist at a company that runs a simple game where players can compete in matches. In the normal course of a match, the match is started when all the players are ready, and it is completed for a given player when they are eliminated or when they win. Players can choose to play as much as they like. When they win, players earn points, which they can spend on cosmetics to personalize their characters. These cosmetics do not grant a competitive advantage against other players.”

Exercise 3, Part 1: How is the game doing?

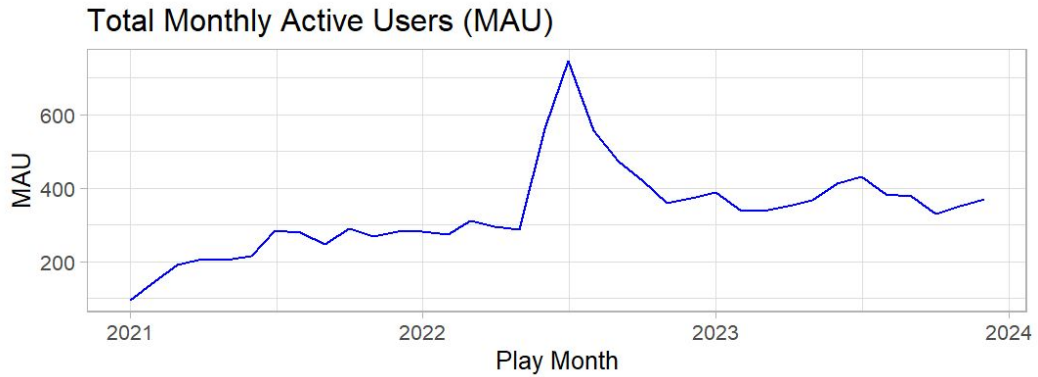
Assumptions/Notes:

- Data provided is based on 100 sampled users per month based on install date and USER_ID.
- As the sample is limited to 100 new users per month, new users are not factored into success metrics.
- As no revenue data is provided, revenue is not factored into success metrics.
- An “active” user is defined by a user that plays or completes a match (more on this in next section).
- Install data is incomplete - analysis window will be truncated to 1/1/2021 - 12/31/2023.
- Datestamps in both files provided are in a common timezone and not local based on the user’s country code.

Exercise 3, Part 1: How is the game doing? (cont.)

User Engagement

- The number of month active users (MAU*) have steadily increased.
- MAU increased through 2021 and began flattening out until an all time high in July 2022 at 746 unique monthly active users.
- After this spike, average MAU remained higher than the pre-spike average.

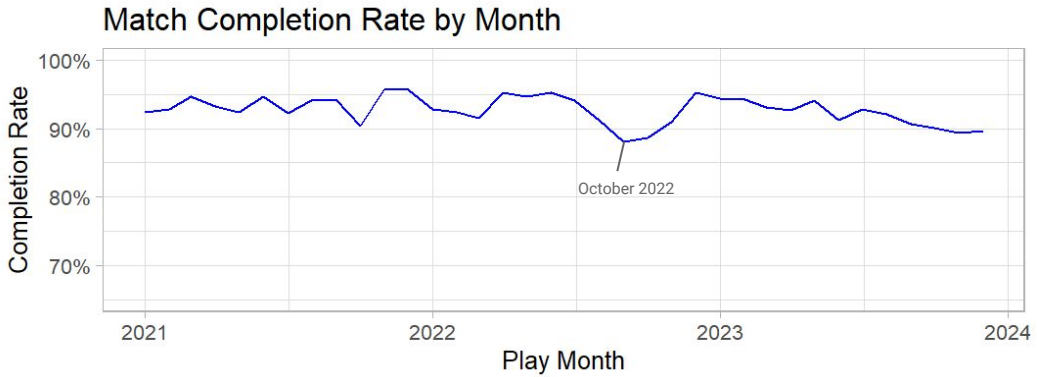


* Active users = A user that has started a match in a given time period
* Monthly Active Users (MAU) = Total unique active users in a given month

Exercise 3, Part 1: How is the game doing? (cont.)

Match Completion Rate

- Match completion rate* dipped in October and November of 2022.
- Completion rate rebounded shortly after but has been steadily declining since.

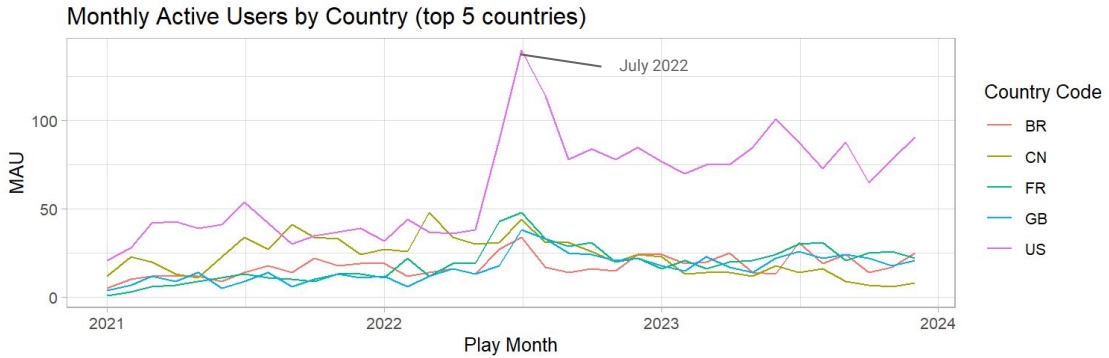


* Completion rate = Total # of matches completed / Total matches started

Exercise 3, Part 1: How is the game doing? (cont.)

Monthly Active Users by Country

- The US remains to dominate in terms of user base.
- Of the 111 countries included in the installs dataset, 20.5% of users are from the US.
- Canada is a bit behind at 8.6% of total active users.



Country	Active Users	% of Total
US	713	20.5%
CN	300	8.6%
BR	181	5.2%
GB	178	5.1%
FR	176	5.0%

* Active users = A user that has started a match in a given time period
* Monthly Active Users (MAU) = Total active users in a given month

Exercise 3, Part 2: Recommendations

Recommendations:

- First and foremost - success metrics need to be determined and communicated. As an analyst, I cannot say whether a game is doing well or not unless I understand what qualifies it as successful. My first recommendation is to sit down and have these conversations up front.

- If MAU is of interest - I would look further into what happened between June and July 2022 to drive so many installs and active users. If there is a lever we can pull - whether it is a marketing campaign, in-game event, partnership, etc., then we should discuss how to best utilize that lever for future planning.

- If match completion rate is a concern - investigate the drop in October/November 2022 to see if there is a way to minimize this concern.

- If churn is a concern a potential next step could be to see if return rate is correlated with win percentage.

Exercise 3, Part 2: How could we improve the data?

Is the quality of the data good?

- Since the data is sampled, the installs data is spotty. Specifically - there is a gap in new users from 5/31/2022 to 6/21/2022.
- Installs data ends at 12/31/2023 while activity data runs through 8/2024 - this leaves us with activity data of only current users in 2024, biasing the user pool and rendering it not useful for this analysis.
- Datestamps should be stored in a more useful format - YYYY/MM/DD or unix timestamp preferred - so that they can be sorted correctly by any language or tool without needing to parse.
- There are two users with NULL country codes in the installs file. These users account for 3 rows in the activities so the NULLs are ignored for the purpose of this analysis.
- There is one user ID that has an install date later than an activity date. Only one activity row is affected, this is ignored.

Are there more things that should be captured?

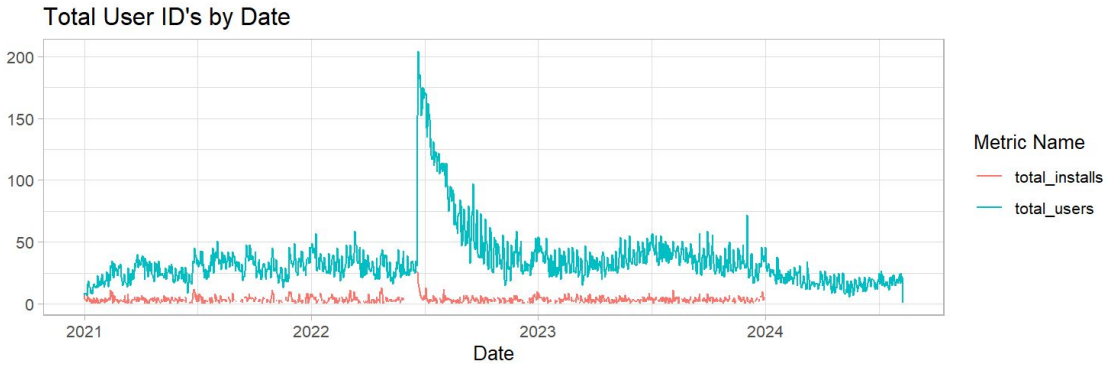
This depends on the points mentions on the previous slide. Additional data (and the work that goes into capturing that data) does not always lead to better insights. That said:

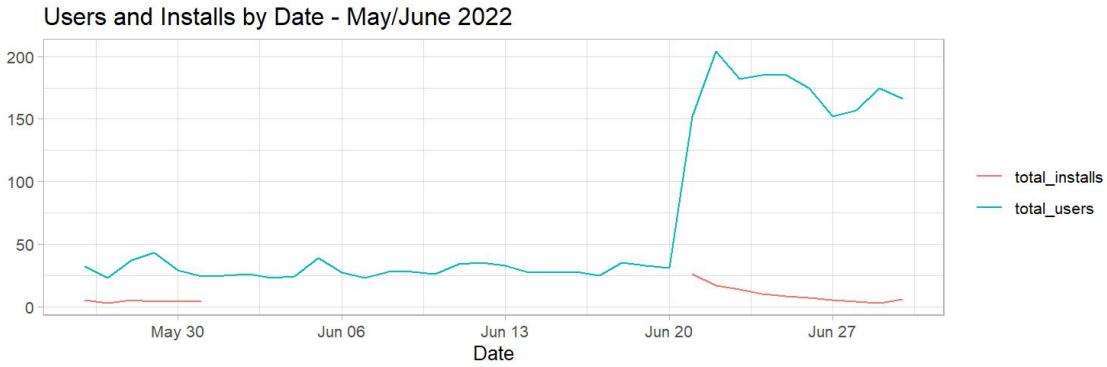
- Cosmetic store data could be important to unearth insights around why users do or do not have motivation to win.
- The reason for a user not completing a match could be helpful for a deeper dive into completion rate. Did the user willfully quit (Pause -> Exit game) or did they disconnect?
- Understanding user churn would require more data. Some points that could be helpful - whether the user plays socially or solo, the device they are playing on (is this game multi–platform?).
- Match length (time) for completed matches, number of players per match, number of winners per match, (and general game context) would help in understanding win rate and how that might impact a user's experience.

End

Appendix

Data gaps:





Appendix

Users vs Active Users vs Active Completion Users