Thank you for taking the time to complete the three exercises below. Please return your response within 7 days of receiving this test.

Instructions/Notes:

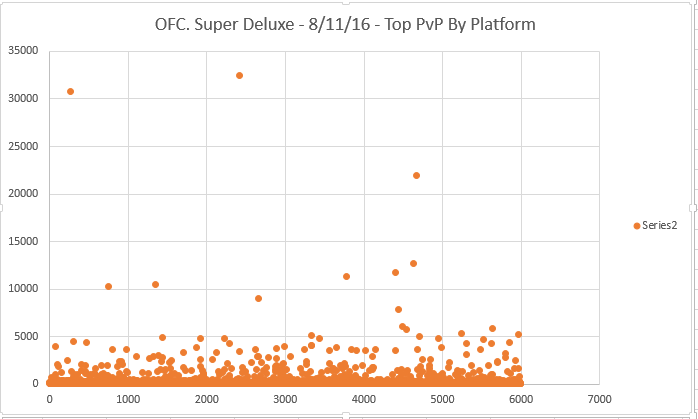
* State your assumptions when you solve each problem.
* Be aware that there may be multiple ways to solve the questions.

**Exercise 1:** 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.

\*\*Datasets\*\*

Exercise1\_webpage\_performance.csv

**Exercise 2:** What’s wrong with these two visualizations and how would you fix them or revisualize them?



1. Poor color scale
2. Values do not encompass entire range
3. Inconsistent data range
4. Order of legend
5. 50% threshold (most important for this dimension of culture) is obscured

**Exercise 3:**

\*\*Context\*\*

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.

You have been asked to summarize the state of the game for non-technical stakeholders, eg. a Product Manager and a Designer. How is the game doing? What are your recommendations for them?

Additionally, how could we improve the data that is currently captured (see datasets attached / described below)? Is the quality of the data good? Are there more things that should be captured?

\*\*Datasets\*\*

You should see the following two datasets in the ZIP file.

\* ` Exercise3\_installs.csv`: This dataset contains 1 row per user and includes the following columns:

\* `USER\_ID`: the unique ID for a given player

\* `INSTALL\_DATE`: the install date for that player

\* `COUNT\_CODE`: the ISO 3166-1 alpha-2 country code for that player

\* ` Exercise3\_activity.csv`: This dataset contains 1 row per user per day and includes the following columns:

\* `USER\_ID`: the unique ID for a given player

\* `PLAY\_DATE`: the date of activity for that player

\* `MATCHES\_STARTED`: the number of matches that player started

\* `MATCHES\_COMPLETED`: the number of matches that player completed

\* `MATCHES\_WON`: the number of matches that player won

\*\*Deliverables\*\*

\* \*\*Deliverable 1:\*\* A summary of how the game is doing & your recommendations for non-technical stakeholders. You can choose whatever format you think is best suited for this. This should be the bulk of your work.

\* \*\*Deliverable 2:\*\* A few notes of how to improve and/or build upon this data.

\* \*\*Deliverable 3:\*\* The code (preferably SQL, Python) or commented workbook (Excel/Google Sheets) you used to run your analysis.