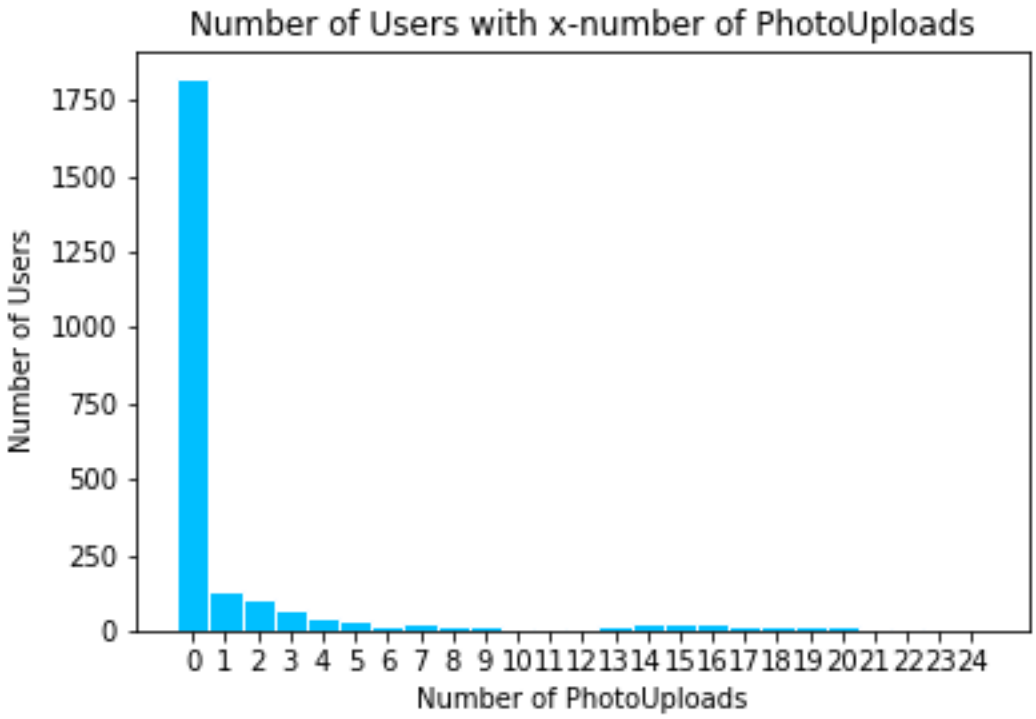
Data Science Challenge Write Up (Results and Conclusions)

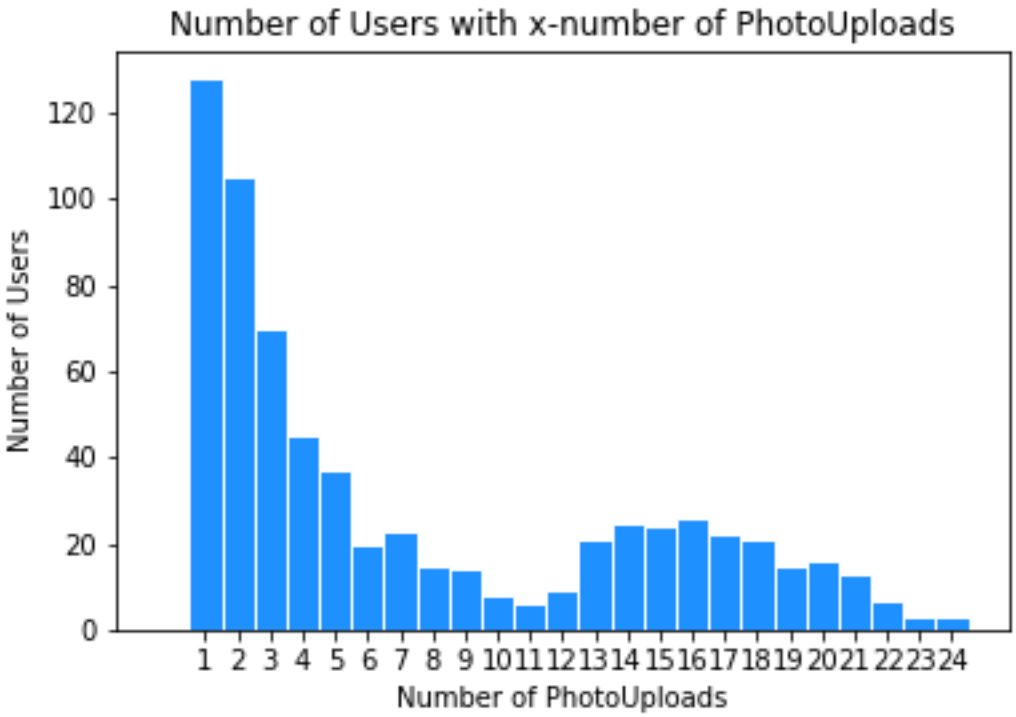
Eric Herrmann

The very first conclusion one can draw from the data is that an overwhelming number of users never upload a photo or interact with the 'PhotoUpload' event. The plot below represents the number of each users in each bin, where each bin is determined by the number of times they triggered the 'PhotoUpload' event.

This histogram indicates that the number of users who are in the bin that indicates that the user never uploaded a photo is by far the largest grouping of users. In fact there are around 15 times as many users in this bin than any other bin. The next largest bin by size represented users who only uploaded a photo once. I considered a pie chart for a follow-up visualization, but the percent of users uploading photos two or more times was far too low of a percentage of the total userbase to make the visualization effective.

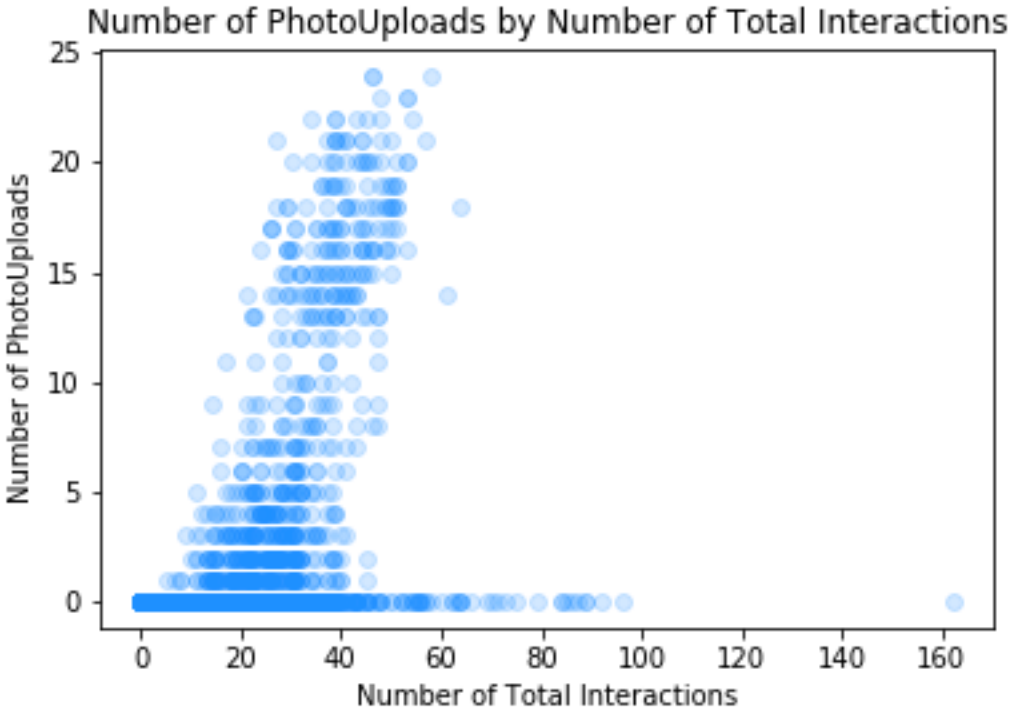
The below diagram excludes the first bin of users, and scales the other bins appropriately to give a better sense of diversity among the remaining userbase.





While this histogram represents a small minority of the userbase, it adds some information as to the diversity of the userbase which uploaded photos at least once. There appears to be an negative relationship between number of users and number of PhotoUploads. However, something of interest happens around x = 15. There are far more users uploading photos between 14 and 18 times than users uploading photos between 8 and 12 times or between 19 and 23+ times. This spike in user density seems to indicate that there is a "sweet spot" where users who upload photos more than a few times settle into this range of uploading photos around 14 to 18 times. This seems to indicate that there are 3 distinct preference groups in the overall userbase: those who never upload a photo, those who upload a few (between 1-5) and those who upload many (between 14-18). This seems to imply that users who would otherwise only upload 6 to 12 photos eventually feel comfortable enough to upload more and more until they reach around 18 and the usefulness of the feature dissipates for them. Overall, this portion of users is still a vast minority of total users.

The third plot below is a scatter graph representing the relation between the number of PhotoUploads by user and the number of total interactions that user made in the app. Total interactions include GuideDownload, GuideSession, ConnectionRequested and PhotoUpload.



The alpha was lowered for this plot to give a better idea of user density. Overall, one can notice that once again there is a large grouping of users at zero PhotoUploads. Some of these users interacted with the app in other ways and still never uploaded a photo. However, there is another interesting relationship present in this plot. For users not in the first grouping of users who never once uploaded a photo, there appears to be a fairly strong linear relationship between the number of total interactions a user made and the number of PhotoUpload interactions the user made. This seems to imply that the more the user interacted with the app as a whole, the more likely they were to upload a photo a larger number of times.

On whole, the data presents several distinct groups of users in regards to their interactions with PhotoUpload. However, the most pervasive group of users are those who never upload a photo altogether.