Cedric Herman
Mentor: Srdjan Santic
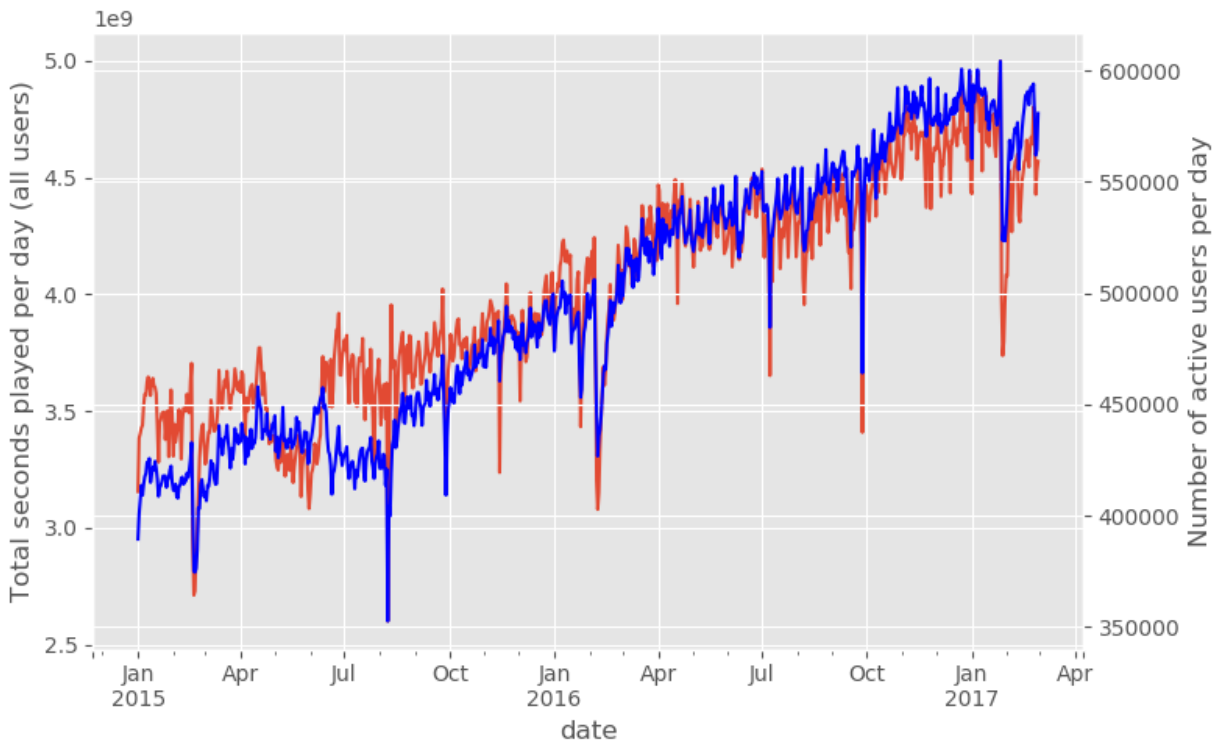Career coach: Allison Matthews

Springboard
January 2018 cohort

Capstone #1: Inferential Statistics

During data exploratory analysis, we made an observation where the daily number of active users seems to correlate with the daily total listening time. Here we will test if this pearson correlation coefficient of 96% is not a result of chance.

Our null hypothesis is that this correlation is actually zero. The alternative hypothesis is that the daily number of users is highly correlated with the daily listening time.

Null hypothesis: number of active users and total listening time are independent.

Alternate hypothesis: there are correlated

We used hacker statistics to compute replicates of this data and compute the probability to obtain a positive correaltion at least as extreme as we observed by chance. This is the definition of the p-value.

Our p-value was so small we couldn't find any instance in our 1 million replicates where we had such a correlation by chance. Therefore we can safely reject the null hypothesis. There is significant evidence that daily number of users and daily listening time are strongly correlated.