Submitted by Krishna Sharma

Part 3: Data Analysis based on 81 users.

Table 1: Summary Statistics

following followers count\_starred public\_repos

count 81.000000 81.000000 81.000000 81.000000

mean 10139.469136 810.753086 27.012346 99.382716

std 44694.607693 2252.516901 8.243928 134.240043

min 0.000000 0.000000 0.000000 0.000000

25% 26.000000 25.000000 30.000000 18.000000

50% 80.000000 86.000000 30.000000 52.000000

75% 280.000000 383.000000 30.000000 122.000000

max 303267.000000 12859.000000 30.000000 849.000000

Table 1 shows the summary statistics for 81 observations here because some of the user don’t have any information. The average number of following is ten thousand and the average number of followers is 810.On an average, the number of public repos is 99.

Similarly, the median value is 80,86 and 52 for following, followers and public repos respectively.

Additionally, only 2.5 percentage of users has no followers and following. Meanwhile, 1.2 percent of users has no public repos.

Examining the correlation between variables

I started with hypothesis that people with high followers tends to have a smaller number of following based on general observation on twitters and Instagram account. So, the first exercise is done in examining the relation between number of followers and following. While doing that, figure 1 shows that there is no clear pattern between followers and following. But interestingly, unlike in twitter, we can see that user with higher followers tends to have higher number of following across GitHub user.

Chart, scatter chart

Description automatically generated

Generally, the user with higher public repos tends to have a greater number of followers. It seems true in our case given we omit some outliers having larger followers. (see Figure 2)

Chart, scatter chart

Description automatically generated

Finally, I have tried including all three variables (following, followers and public repos) in one figure to re-iterate the findings discussed above.

Chart, scatter chart

Description automatically generated