As we increase the number of subsets of data we are testing, the variance we see in which k-value is best will increase. Say we have a data set and we are testing which k value gives the best prediction from k=2 to k=6, and that k=6 is the true best value. Let's say that 60% of the subsets of our data find k=6 to produce the best prediction. The other 40% of the time, the other k values are equally likely, each being best for 10% of the subsets. Let's look closer at k=2. If we run the function on one random subset, there is a 10% chance that 2 will be returned as the best k value. However, if we run the function on two subsets, the chance that one of these subsets has k=2 as the best value increases to 19%. Run the function on 7 subsets and it's more likely than not that k=2 is the best value for one of these subsets. With this being the case, we can understand why the variance in best k value increases as the number of data subsets we are using increases. It is because as the number of trials we are conducting increases, the likelihood that a sunset has a best k value that is different from the best k value for the full scope of the data increases, thus increasing the diversity in best k value results, increasing the variance between these k values. This is a good example of p-hacking because it could be used to find significant results where significant results do not really exist. Say for the scenario above, we were involved in a competition where k=2 happens to explain the test set best. We could show that the model we used generated k=2 as the best k value; however, doing so is technically manipulating our choice of inputs so that we get a statistically significant result where no statistically significant result can really be found.