Cliques in r/place

The main goal of this analysis was to find "cliques", or groups of users exhibiting similar behavior with their participation in r/place. Through K-Means clustering, I tasked myself with determining whether these groups could be placed into interpretable clusters, such as "casual contributors", long-term enthusiasts, or "rapid" superhumans (or even "bots") based purely on activity patterns, timing, and spatial behavior.

For each user, a feature vector captured three different behaviors that may contribute to what cluster they would end up in.

Activity/Intensity:
total number of pixel placements made by the user throughout the entire event

- `active_duration`: calculate `max(t_ms) - min(t_ms)` to find how long they participated for (were these users just participating in a fun drive-by style, or were they in this for the long run?) "Capture whether users participated briefly or remained over a long portion of the event"

Skillset

- `median_window_ms`: median inter-event time for every user
- `fast_ratio_p01`: fraction of the user's windows below a predetermined "fast" threshold

Creativity:

- `unique_pixels`: what is the number of distinct (x, y) objects did the user touch throughout their participation
- `spatial_spread`: compute `(max_x - min_x + 1) * (max_y - min_y + 1)` to create a bounding box of where each user worked (small bounding box could hint at the idea that a user focused on their craft, while a large bounding box could imply that a user didn't emphasize areas of the canvas but worked across many regions of the canvas)

Because these features were on different scales (total events could span from 1 to a thousands, whereas the fast ratio is bounded between 0 and 1), a couple things were done to the data to make it suitable for K-Means clustering: log-transforming various features as well as using standardizing all of them to prevent feature domination.

K-Means clustering was then applied to standardized feature matrix. After experimenting with various values of k, 4 was selected to produce reasonable sized and interpretable groups. This resulted in four clusters being created; three were large, human-scale clusters, whereas the last one was extremely small and unnatural.

cluster --- i32	n --- u32	mean_events --- f64	mean_duration --- f64	mean_median_window --- f64	mean_fast_ratio --- f64	mean_unique_pixels --- f64	mean_spread --- f64
0	3642495	37.151423	136901.530915	1.0070e6	0.013639	35.011509	2.1111e6
1	2411117	5.043265	115861.193415	3.4352e7	0.000931	4.949214	914514.054618
3	1899822	4.559747	6469.210369	1.1753e6	0.00374	4.391527	400754.182253
2	87231	21.392899	33004.60929	1.7341e6	0.457397	18.130195	643204.865392

Cluster 0, the largest of them, has the longest active duration and moderate inter-event window (about 17 minutes). The mean unique pixels placed by users in this cluster is 35, and there is large spatial spread present; the users in this cluster can be described as committed human contributors who actively participated for a long time and worked across large areas

Cluster 1 has the lowest activity of any other cluster. With only 5 events, sparse actions (extremely large inter-event time of about an hour), and only 5 unique pixels with a very small spatial spread, these users can best be characterized as “touristy”; they contributed once or twice each and probably only placed pixels for fun, not because they were taking the event seriously or were committed.

Cluster 3 exhibited similar behavior to Cluster 1, with low average events per user and small spatial spread. The main difference is the slightly faster timing than cluster 1, which may be an indication that users in this cluster joined briefly to help with a small goal in mind and then left.

The most interesting cluster in my opinion is Cluster 2 with highly suspicious numbers and almost automated behavior; although activity is not suspiciously high at an average of 21 events, the “fast ratio” is near 0.50, indicating almost half of the actions of the users in this cluster are “fast”. Additionally, there are not as many unique pixels as well as those un Cluster 0. These factors do not mean the users in this cluster are bots, but could hint at the idea.