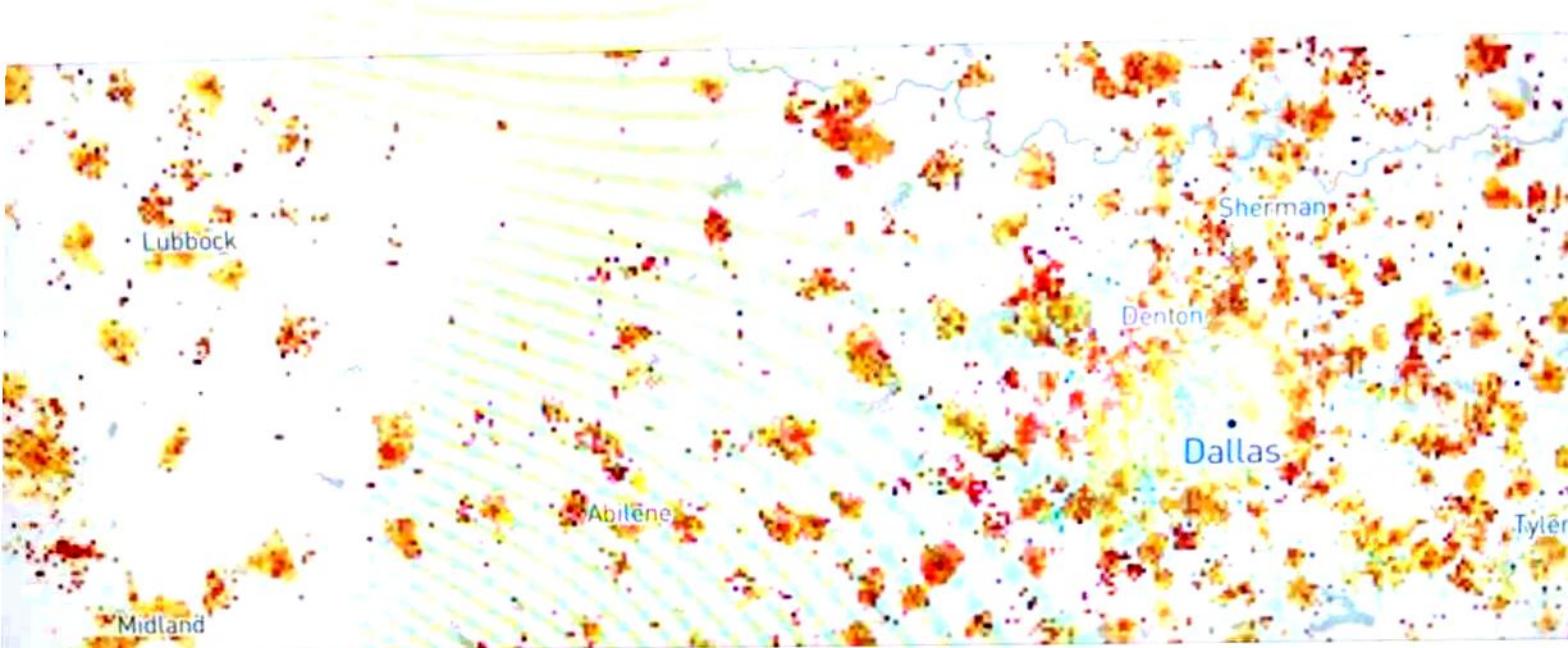


Likelihood of being in the Heavy/Medium user segments (Luddites are filtered).
Detail.



Are local groups bad for engagement?

TUESDAY, OCTOBER 15, 2019 · READING TIME: 7 MINUTES

REDACTED FOR CONGRESS

In this note I present four loosely connected results stemming from one analysis:

1. Engagement with groups has clear geographical patterns
2. In the US - city dwellers are less engaged in groups
3. Overall, in the US, demographics is a good predictor of group engagement
4. The number of local groups near a person is negatively correlated with engagement

In this note I present four loosely connected results stemming from one analysis:

1. Engagement with groups has clear geographical patterns
2. In the US - city dwellers are less engaged in groups
3. Overall, in the US, demographics is a good predictor of group engagement
4. The number of local groups near a person is negatively correlated with engagement

Engagement with groups has clear geographical patterns

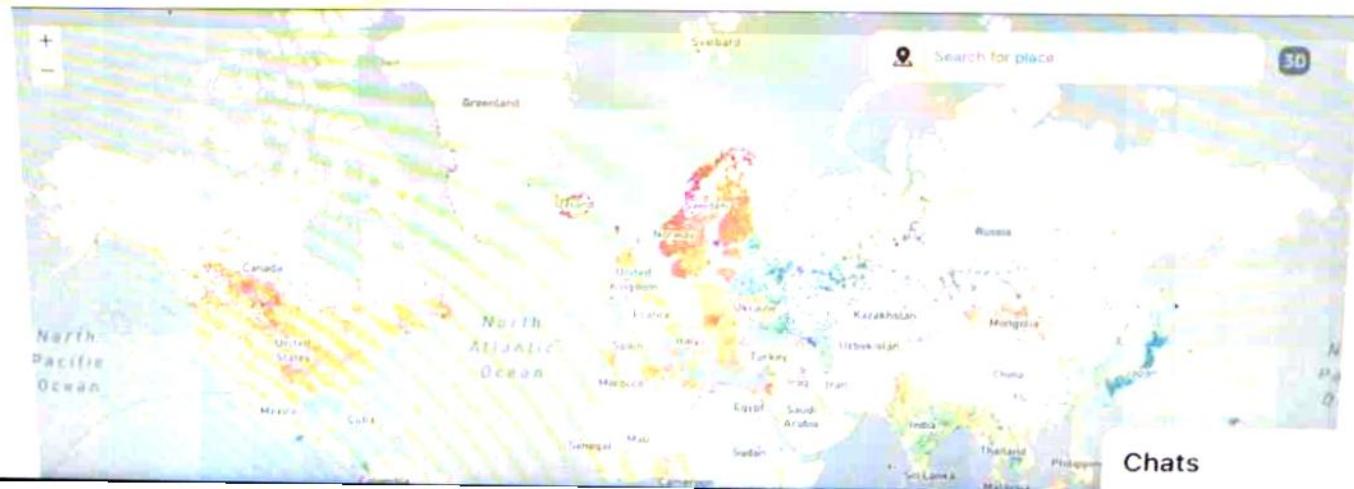
I've started my analysis with a goal of finding sub-populations underserved by the current groups ecosystem. My thinking was that by finding such clusters of low engagement, we might be able to create or promote the creation of groups dedicated to those populations. Hopefully, such a concentrated effort will be more effective in identifying inventory-gaps.

The groups connect team segmented the population of FB users in the following way:

1. **Luddites** - People that are not active in groups, and are unlikely to become active even if we tried our hardest. Technically - people that are not members of any active group, having $l7 \leq 2$ and either don't have a profile picture or have joined Facebook more than a year ago.
2. **No** - People that are not members of any active group (and are not Luddites)
3. **Light** group users- People that are members of 1 to 10 active groups, and their group $l28$ is no more than 7 days.

1. **Luddites** - People that are not active in groups, and are unlikely to become active even if we tried our hardest. Technically - people that are not members of any active group, having $l7 \leq 2$ and either don't have a profile picture or have joined Facebook more than a year ago.
2. **No** - People that are not members of any active group (and are not Luddites)
3. **Light** group users - People that are members of 1 to 10 active groups, and their group $l28$ is no more than 7 days.
4. **Medium** group users - Best defined as people that are not any of the other segments
5. **Heavy** group users - People that are members of at least one active group, and their group $l28$ is at least 14 days.

With that definition, I filtered out all Luddites, and plotted in the next map the other segments. The more red the color, the more likely that the people living in that "pixel" to be in either the Medium or Heavy segments.





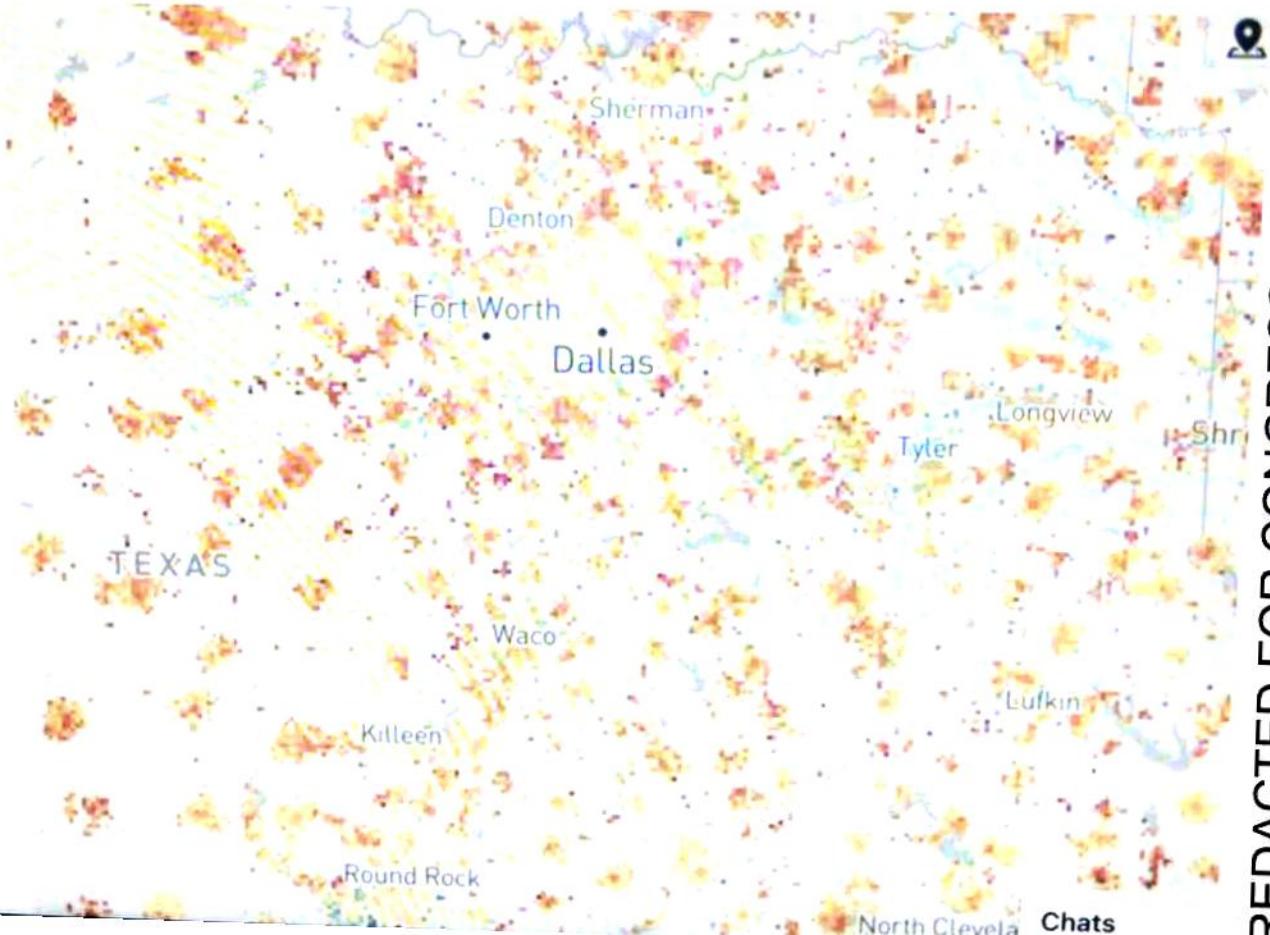
Likelihood of being in the Heavy/Medium user segments (Luddites are filtered)

It's not super surprising to see that groups-engagement, like overall FB-engagement is very much dependent on geography. Still, it's interesting to note that the effort led by the groups connect team, to push NO/Light users into more engaged segments, means as a matter of fact, focusing on some specific countries.

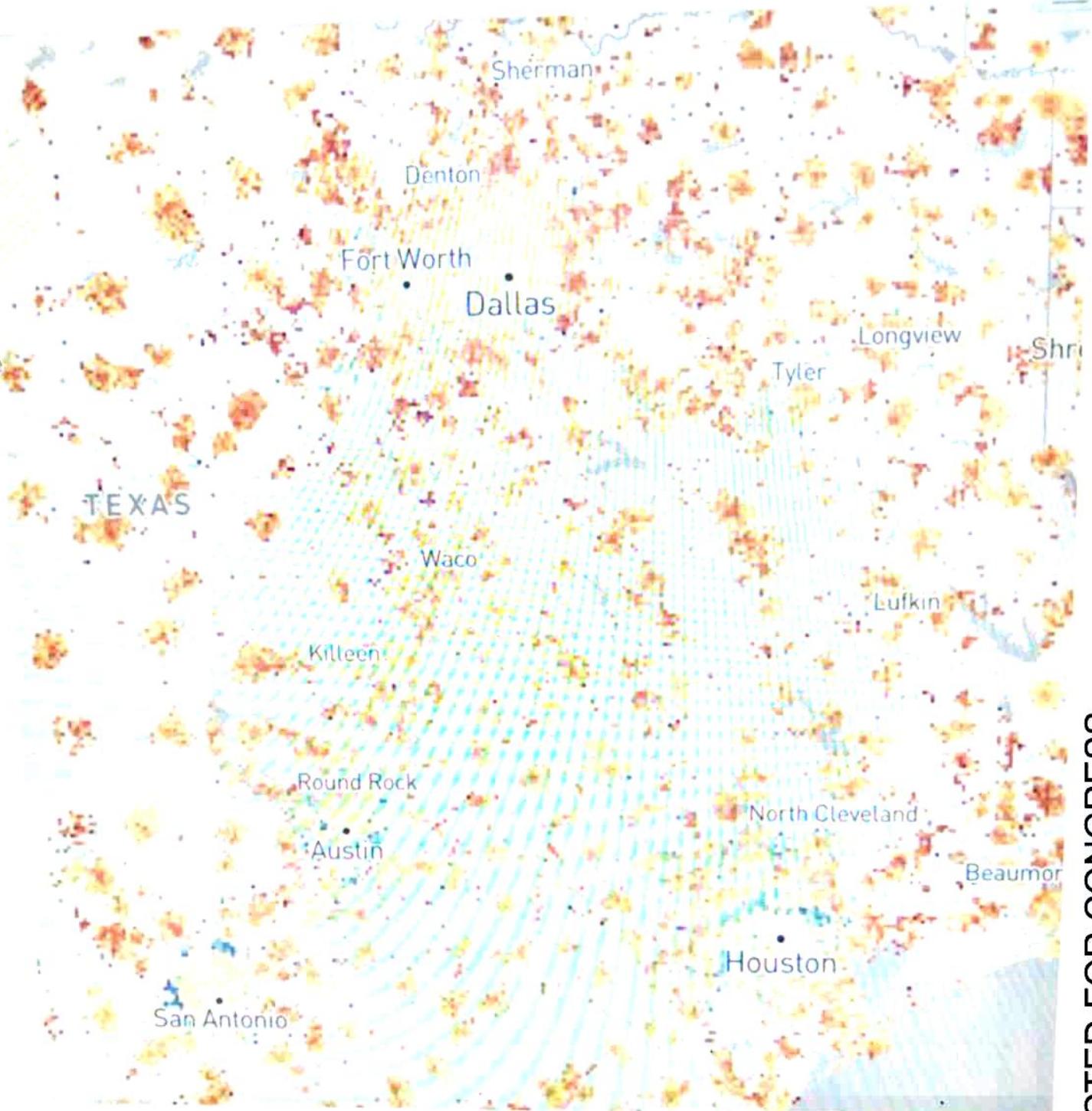
In the US - city dwellers are less engaged in groups

In the US - city dwellers are less engaged in groups

Focusing on the US, one thing is apparent, even in the map above - people living near the coast are less engaged with groups in comparison to people living in the heartland. Zooming in on Texas reveals another interesting pattern:

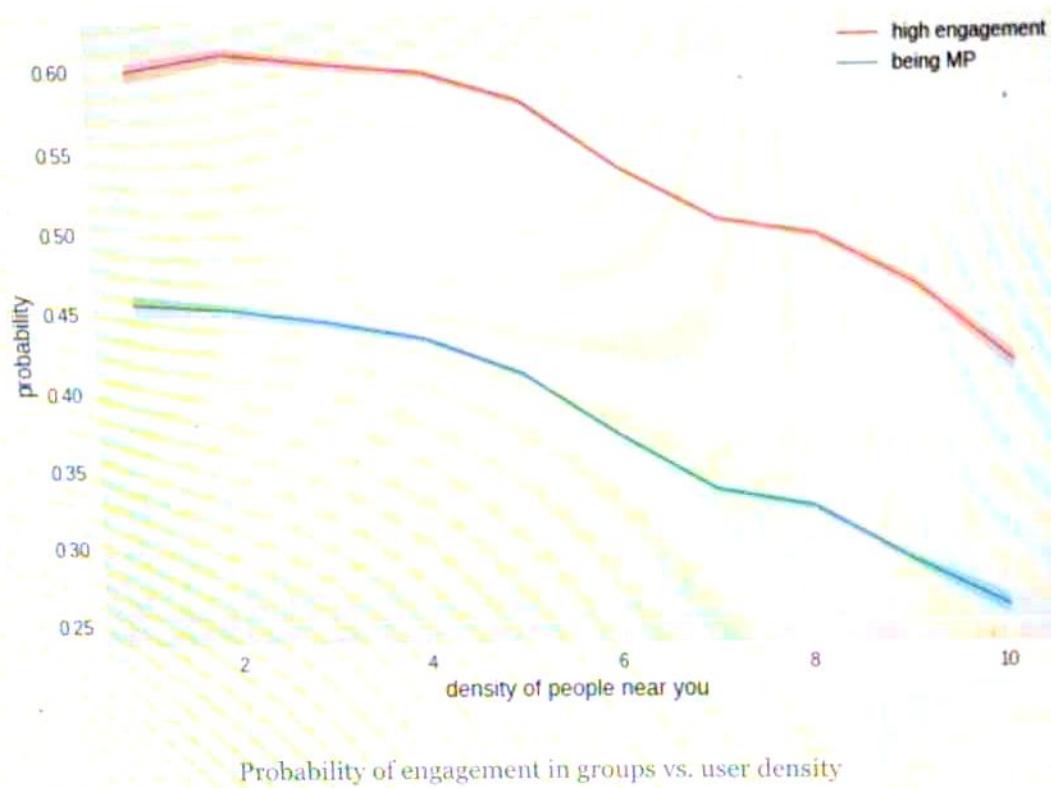


REDACTED FOR CONGRESS



Likelihood of being in the Heavy/Medium user segments (Luddites are filtered).
Detail.

The big population areas are less engaged with groups than the peripheries. We can see it in a more objective way in the following graph, plotting likelihood of engagement (and likelihood of being MP) as a function of user density:



This brings up some questions and a few possible explanations. [REDACTED] for example is working on a more substantial analysis linking group engagement with social capital. I won't try to fully explain why people in big cities are less engaged with groups. I'll note though that user density is also negatively correlated with overall l28 at Facebook and likelihood of being exposed to GYSJ, though the correlation is not as strong as in the above graph.

This brings up some questions and a few possible explanations. █ for example is working on a more substantial analysis linking group engagement with social capital. I won't try to fully explain why people in big cities are less engaged with groups. I'll note though that user density is also negatively correlated with overall l28 at Facebook and likelihood of being exposed to GYSJ, though the correlation is not as strong as in the above graph.

Overall, in the US, demographics is a good predictor of group engagement

Given that user-density is correlated with group engagement, a natural follow-up question is how much can we predict group-engagement given only user demographics features. Or more succinctly - **can we predict group engagement without using any group related features?**

Surprisingly the answer for US based users is yes. **Given only user demographics features, group engagement and MP status can be predicted with high accuracy.** This result (which was known to some people) is interesting on its own - it suggests that the efficacy of our efforts to make people more engaged, or to create more meaningful people, is somewhat limited by demographics.

Specifically we define two classification tasks:

1. Is a US user in the High/Medium user segments (versus the No and Light ones, ignoring Luddites)?
2. Does a US user have a meaningful connection with any group - that is, are they meaningful people.

Specifically we define two classification tasks:

1. Is a US user in the High/Medium user segments (versus the No and Light ones, ignoring Luddites)?
2. Does a US user have a meaningful connection with any group - that is, are they meaningful people.

As features I use the following - l1, l7, l28, number of friends, nearby user density, facebook age, age, gender and mobile OS. Note again that I don't include any features related to groups, like "number of groups that is user is member of". I fit a GBDT for the two tasks, without any fine tuning or metaparameter search (just sklearn's defaults).

When predicting user segment, the model gets AUC=0.81, meaning given two random users, one is highly engaged with groups, one is not, we'll be able to identify which is which 81% of the times. The accuracy is 73%, while a constant model would get 51%. L28 and friend count are the most important features.

Predicting MP is a bit harder, but we still get AUC=0.79. The accuracy of the model is 74%, while a constant model would get a 67%.

The number of local groups near a person is negatively correlated with engagement

Finally, let's look at whether more local groups is better for engagement. Obviously, given that user-density is negatively correlated with engagement, we can expect that density is negatively correlated with engagement. That is, in his cities, there are Chats

The number of local groups near a person is negatively correlated with engagement

Finally, let's look at whether more local groups is better for engagement. Obviously, given that user-density is negatively correlated with engagement, we can expect that local group density is negatively correlated with engagement. That is, in big cities, there are many local groups, but users are less engaged.

To control for user-density and other demographics factors, I use the following procedure:

1. Take the classifier described in the last section and for a new set of users predict "baseline" engagement probability, without knowledge of the number of nearby local groups.
2. Train a new classifier (logistic regression), using the baseline probability from (1) and the (log) number of active local groups near the user, trying to predict engagement.
3. Observe the weight the model gives to the number of active local groups.

[technical note - by active local groups I mean public or closed, non BSG, groups whose p75 distance from their centroid is less than 25KM, and having at least 3 distinct users posting or commenting in the last week. We're counting how many such groups are within 25KM of the user]

In both the user-segment task, and in the MP prediction task, the number of nearby local groups is actually having a negative weight, which is significantly below 0. Obviously, I'm not claiming that more local groups cause less engagement with groups overall - this is just a correlation. But, I am now skeptical of any efforts to increase the number of 1 Chats

In both the user-segment task, and in the MP prediction task, the number of nearby local groups is actually having a negative weight, which is significantly below 0. Obviously, I'm not claiming that more local groups cause less engagement with groups overall - this is just a correlation. But, I am now skeptical of any efforts to increase the number of local groups in hopes of increasing engagement. Following this analysis, **I believe that the number of local groups available to the user is uncorrelated with the user's engagement.**

	coef	std err	z	P> z	[0.025	0.975]
intercept	-2.3925	0.019	-124.447	0.000	-2.430	-2.355
baseline probability	5.1463	0.018	291.186	0.000	5.112	5.181
log(nearby local groups)	-0.0302	0.002	-12.859	0.000	-0.035	-0.026

Linear model fitted to predict being in the heavy/medium user segments, showing that local groups has a significantly negative coefficient with engagement

Closing Notes

I started this analysis hoping that I will be able to identify unengaged sub-populations, that will benefit from us creating more local groups dedicated for them. I conclude with a somber realization that this might be too optimistic, and being engaged with groups is not as strongly correlated, at least in the US, with available inventory.

CC :

for inventory - is our GYSC approach, looking to Chats

Closing Notes

I started this analysis hoping that I will be able to identify unengaged sub-populations, that will benefit from us creating more local groups dedicated for them. I conclude with a somber realization that this might be too optimistic, and being engaged with groups is not as strongly correlated, at least in the US, with available inventory.

CC :

- [REDACTED] for inventory - is our GYSC approach, looking to create local groups a bit too optimistic?
- [REDACTED] - on whether connecting more users from the low-engagement segments limited by demographics.

Thanks [REDACTED] for the discussions around this note.

My Python notebook with the results listed above is at n156946



42 Comments 8 Shares

Like

Comment

Share

Save

View 10 more comments

[REDACTED] really interesting insights! Couple of thoughts

- on user density being negatively correlated with engagement: similar to [REDACTED]'s comm these communities are engaging in real life more easily, meaning that they don't have to enga Chats

Like · Reply · 1y

██████████ Would love to hear from you █████ and █████ why people in big cities are less engaged with groups! (even an initial set of hypotheses which still need proving out)

Like · Reply · 1y · Edited



██████████ Does this relate at all to the Facebook app in flyover states vs. coastal correlation with sessions/TS/DAP?

Like · Reply · 1y



██████████ "flyover states" is a true coastal elite phrase



Like · Reply · 1y

██████████ My hypothesis is that people in cities have compelling offline alternatives to whatever value FB Groups provide. But that seems a bit simplistic.

Like · Reply · 1y



██████████ I was actually thinking the same thing, but also maybe that other online options (Meetup, Neighborly, etc) work better for higher population density areas.

██████████ has any of your research on competitors looked into City vs. Rural usage that would validate/invalidate this hypothesis? If this is true, it opens up an entirely new area of work 😊

Like · Reply · 1y



██████████ My intuition is that other online options such as Meetup, Neighborly, etc. are too small to account for the majority of the difference. If I recall correctly, the population penetration for those products is quite low and the time spent per DAP is also low. I haven't tried quantifying that though. Would be super interesting to see the total TS for competing online products vs the FB Group delta between city vs non-city.

Like · Reply · 1y



██████████ Agree w/ █████ that competitors aren't driving the difference, but still helpful to look at. What we see is that U.S. Nextdoor users are much less likely to be rural vs. FB overall in U.S. (<https://fburl.com/daiquery/5i97hv8h>). We've also seen that overall FB app (or █████; I don't have Groups) usage (L7, sessions) is stronger on average when users are also using Nextdoor. Again, my interpretation is that it's not about a specific competitor, but does speaks to our having really strong PMF in less dense areas in U.S.

Like · Reply · 1y



██████████ nextdoor u.s. pop'l density sample for those interested

Chats

Like · Reply · 1y

[REDACTED] Agree w/ [REDACTED] that competitors aren't driving the difference, but still helpful to look at. What we see is that U.S. Nextdoor users are much less likely to be rural vs. FB overall in U.S. (<https://fburl.com/daiquery/5i97hv8h>). We've also seen that overall FB app (or [REDACTED], I don't have Groups) usage (L7, sessions) is stronger on average when users are also using Nextdoor. Again, my interpretation is that it's not about a specific competitor, but does speaks to our having really strong PMF in less dense areas in U.S.

1

Like · Reply · 1y

[REDACTED] nextdoor u.s. pop'in density sample for those interested

Nextdoor # U.S. users by geo population

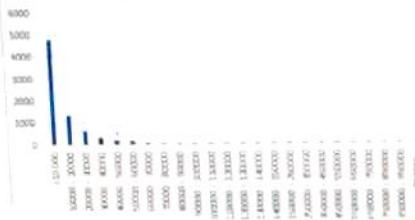


2

Like · Reply · 1y

[REDACTED] fb u.s. pop'in density sample for those interested. cc Steve who was also asking about this

Facebook # U.S. users by city population



2

Like · Reply · 1y

Write a reply...

[REDACTED] cc [REDACTED]

Like · Reply · 1y

REDACTED FOR CONGRESS

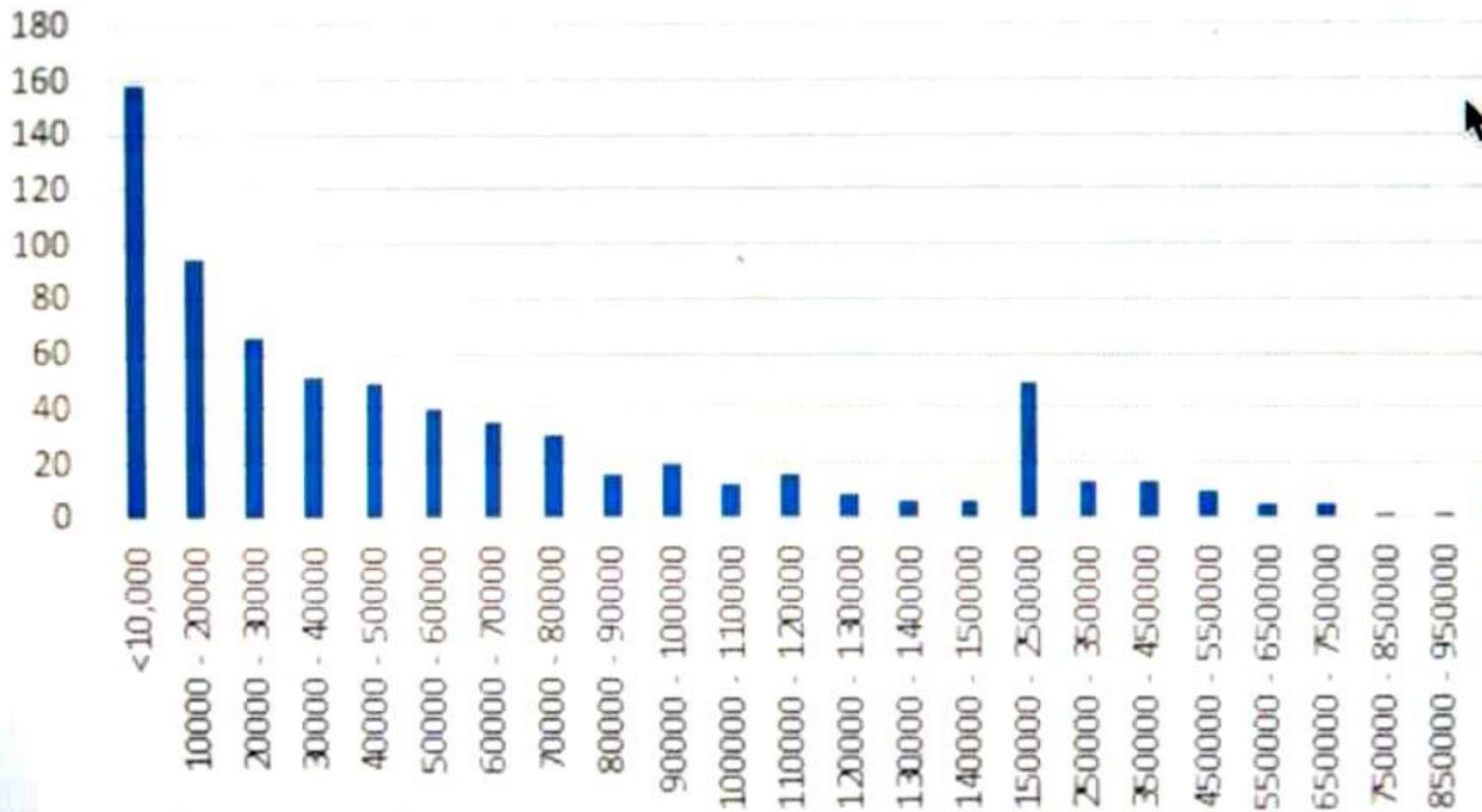
[REDACTED] , cc [REDACTED] on demographics being a good predictor of a group (popular groups, category UIP)

Chats

Israel could this suggest that our recommendations are currently biased at engaging only certain c

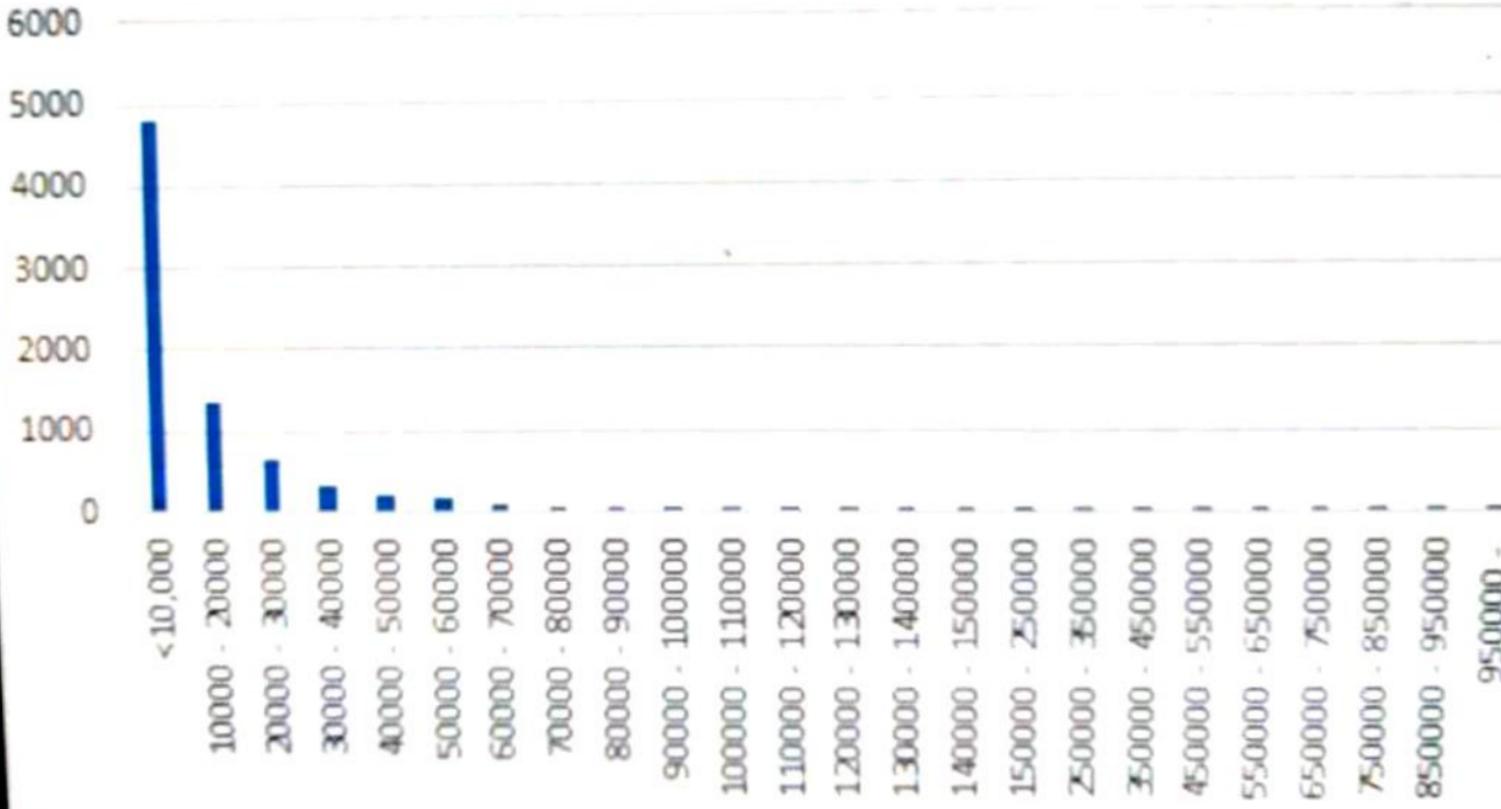
Save

Nextdoor # U.S. users by geo population



REDACTED FOR CONGRESS

Facebook # U.S. users by city population



REDACTED FOR CONGRESS

[REDACTED] cc [REDACTED]

Like · Reply · 1y



[REDACTED] cc [REDACTED] on demographics being a good predictor of a group (popular groups, category UIP)

[REDACTED] could this suggest that our recommendations are currently biased at engaging only certain demographics? Is it a chicken/egg problem?

Like · Reply · 1y Edited



[REDACTED] they are. but not sure how is it a chicken and an egg scenario - if people become less engaged, they are more likely (with recent changes) to be targeted by our recommendations. So the feedback loop should be broken?

Like · Reply · 1y



[REDACTED] for # of local groups near a person as negatively correlated with engagement

Like · Reply · 1y



[REDACTED] Interesting reading. I assume by saying looking to create more local group is too optimistic, you mean # local contributes little to none(or even negative) to likelihood of engagement. I wonder if it's possible # local groups could actually attract more users to group and then absolute number of engaged people would increase given lower likelihood of engagement

Like · Reply · 1y



[REDACTED] Good point!

When doing the same procedure just to predict NO users, there's still a positive correlation between number of local groups and being a NO. Since NO are not necessarily users that know about groups, I think your idea doesn't explain everything we see here.

Like · Reply · 1y



[REDACTED] Super interesting analysis! I wonder if the reason for this effect is that people in cities have a different demand for different types of groups. Eg. people in rural areas might have a high demand for general neighborhood groups (we know this from research!), while people in cities have a higher demand for organization, interest, and identity x local groups.

I'd be interested to see this analysis broken down by group type to confirm this (I don't think we have a good existing model to predict if a group is local interest, identity, organization, culture, or neighborhood, which might make it tough). I'd expect to see that rural areas have high demand for neighborhood groups, while cities have high unmet demand for local x something groups. (cc [REDACTED] for this hypothesis.)

Second hypothesis: people in cities use other apps instead of Groups, like Meetup, Eventbrite, and Reddit. (cc [REDACTED] for this hypothesis.)

Like · Reply · 1y

Chats

[REDACTED] Super interesting analysis! I wonder if the reason for this effect is that people in cities have a different demand for different types of groups. Eg. people in rural areas might have a high demand for general neighborhood groups (we know this from research!), while people in cities have a higher demand for organization, interest, and identity x local groups.

I'd be interested to see this analysis broken down by group type to confirm this (I don't think we have a good existing model to predict if a group is local interest, identity, organization, culture, or neighborhood, which might make it tough). I'd expect to see that rural areas have high demand for neighborhood groups, while cities have high unmet demand for local x something groups. (cc [REDACTED] for this hypothesis.)

Second hypothesis: people in cities use other apps instead of Groups, like Meetup, Eventbrite, and Reddit. (cc [REDACTED] for this hypothesis.)

Like · Reply · 1y

3

▲ Hide 15 Replies

[REDACTED] I'm actually dying to see how such a map looks for Instagram 😊

Like · Reply · 1y

[REDACTED] might have some thoughts on that one?

Like · Reply · 1y

[REDACTED] what do you mean by "map for Instagram"? Because Instagram doesn't have the analogous Groups product... If we were just looking at DAP or Time Spent IG would surely do better in the coastal areas.

Like · Reply · 1y

[REDACTED] Example: <https://fb.workplace.com/photo.php?fbid=519981488475233&set=p.519981488475233&type=3&theater>



Like · Reply · 1y · Edited

1

[REDACTED] yeah, I meant some arbitrary engagement metric. The link you posted is broken for me though.

Like · Reply · 1y

Chats

[REDACTED] Super interesting analysis! I wonder if the reason for this effect is that people in cities have a different demand for different types of groups. Eg. people in rural areas might have a high demand for general neighborhood groups (we know this from research!), while people in cities have a higher demand for organization, interest, and identity x local groups.

I'd be interested to see this analysis broken down by group type to confirm this (I don't think we have a good existing model to predict if a group is local interest, identity, organization, culture, or neighborhood, which might make it tough). I'd expect to see that rural areas have high demand for neighborhood groups, while cities have high unmet demand for local x something groups. (cc [REDACTED] for this hypothesis.)

Second hypothesis: people in cities use other apps instead of Groups, like Meetup, Eventbrite, and Reddit. (cc [REDACTED] for this hypothesis.)

Like · Reply · 1y

3

▲ Hide 15 Replies

[REDACTED] [REDACTED] I'm actually dying to see how such a map looks for Instagram 😊

Like · Reply · 1y

[REDACTED] [REDACTED] might have some thoughts on that one?

Like · Reply · 1y

[REDACTED] [REDACTED] what do you mean by "map for Instagram"? Because Instagram doesn't have the analogous Groups product... If we were just looking at DAP or Time Spent IG would surely do better in the coastal areas.

Like · Reply · 1y

[REDACTED] [REDACTED] (PM) Example: <https://fb.workplace.com/photo.php?fbid=519981488475233&set=p.519981488475233&type=3&theater>



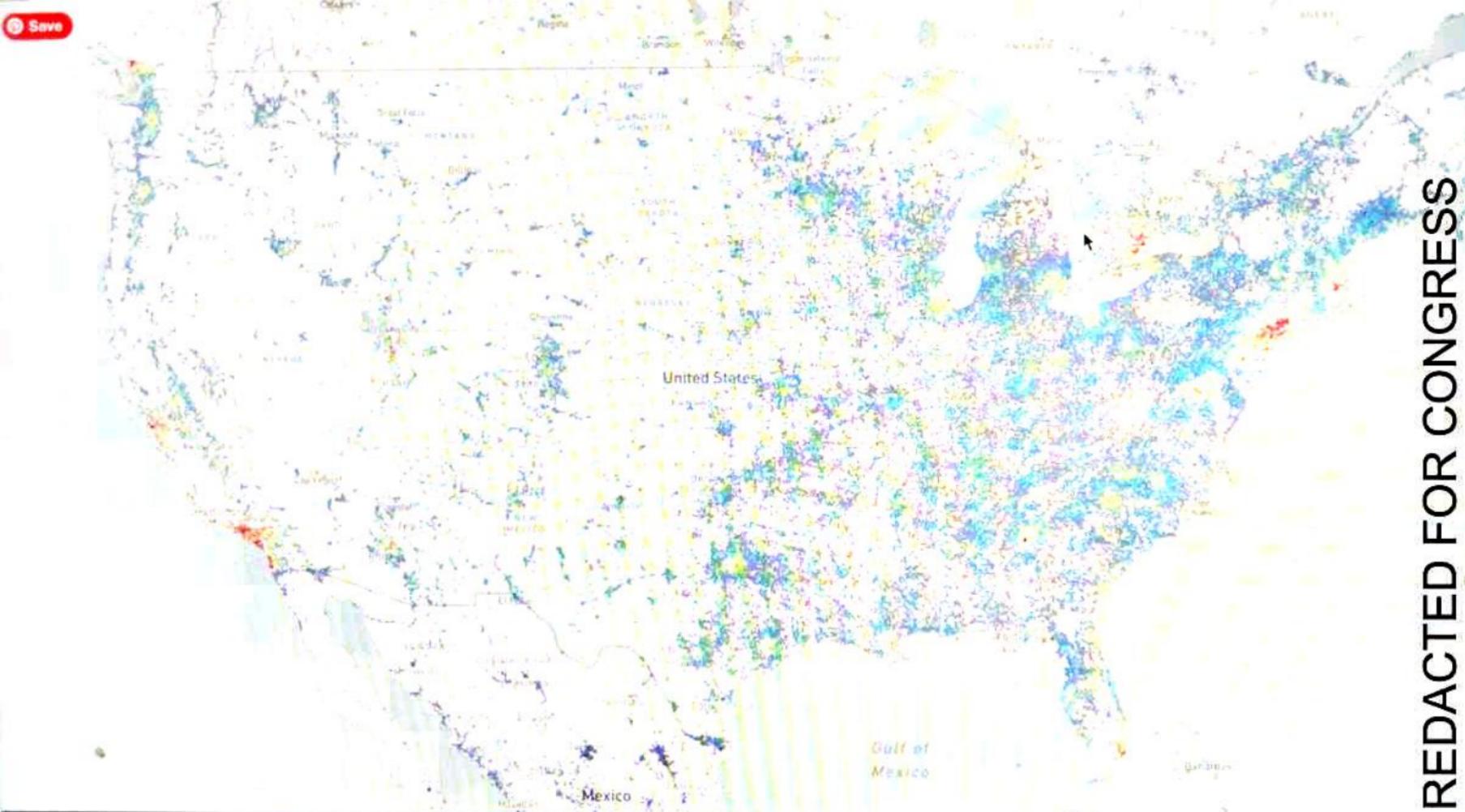
Like · Reply · 1y · Edited

1

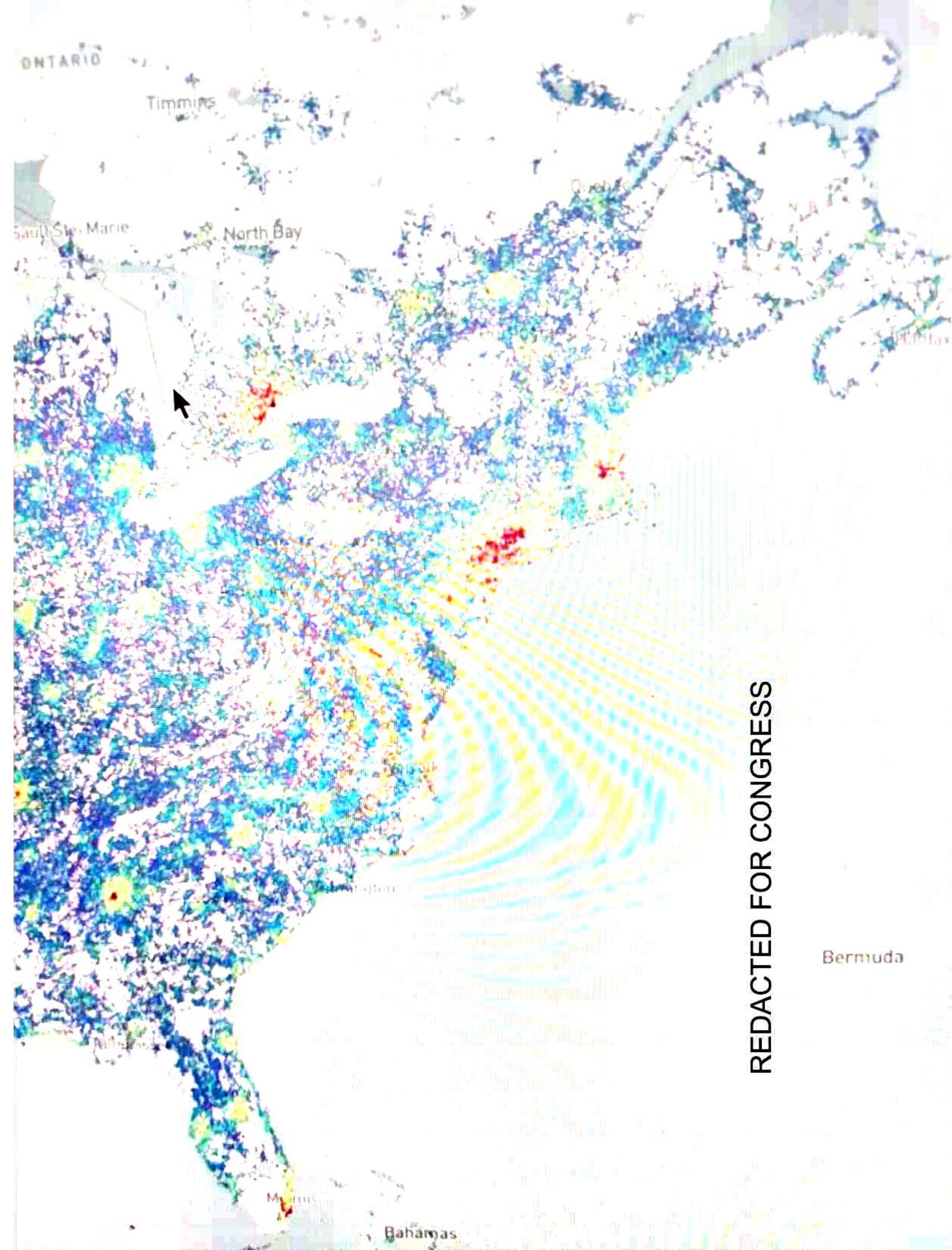
[REDACTED] [REDACTED] yeah, I meant some arbitrary engagement metric. The link you posted is broken for me though.

Like · Reply · 1y

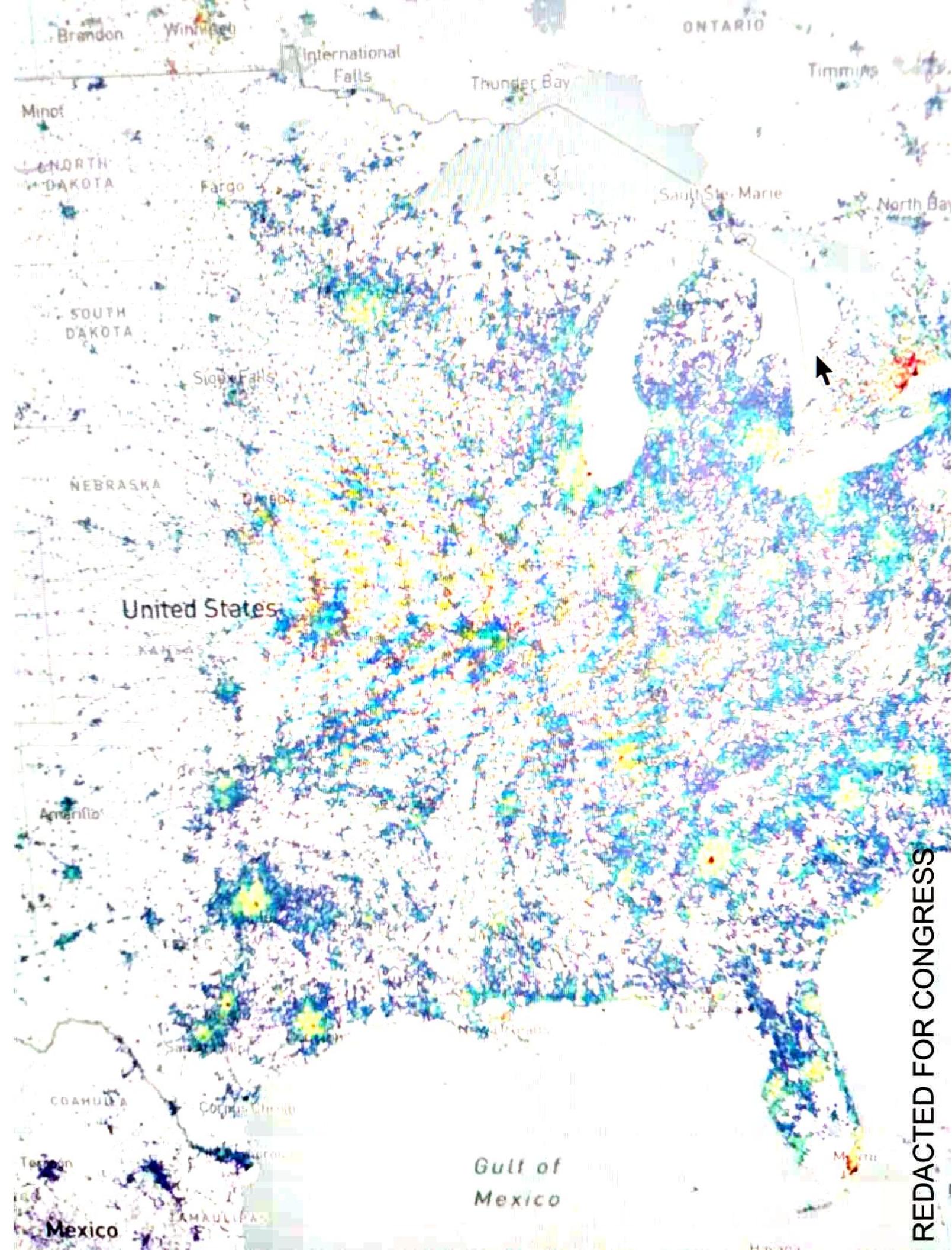
Chats



REDACTED FOR CONGRESS

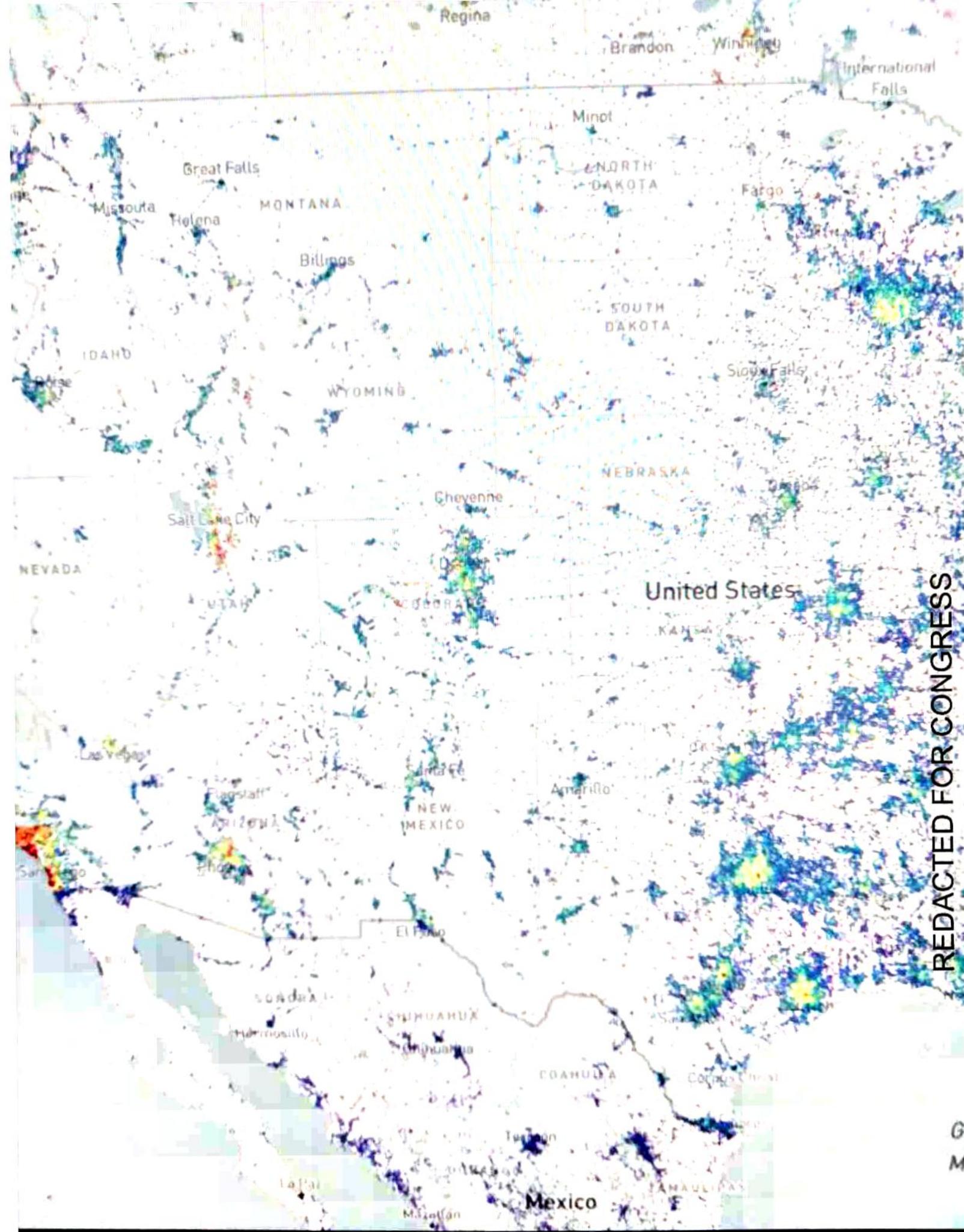


REDACTED FOR CONGRESS



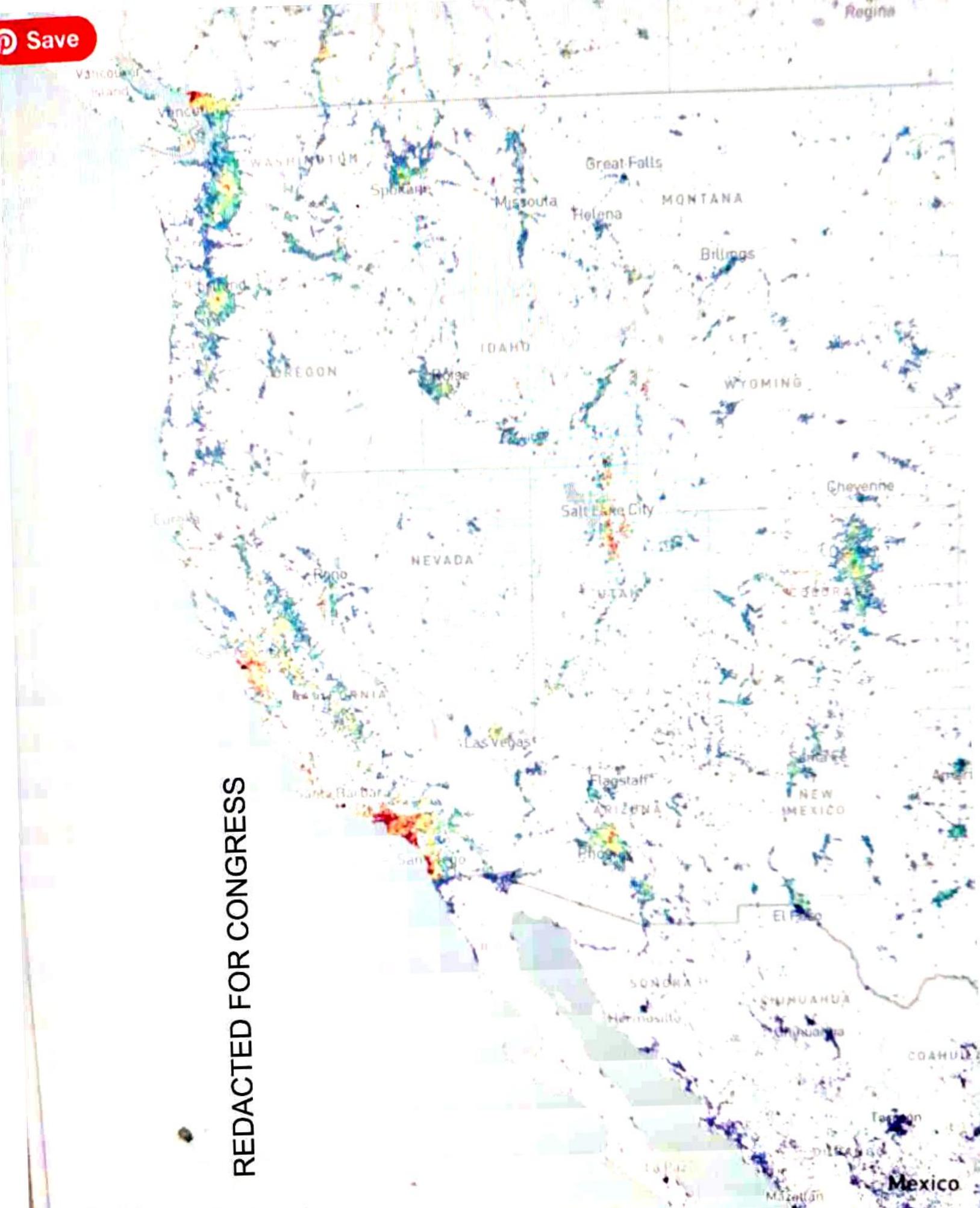
REDACTED FOR CONGRESS

REDACTED FOR CONGRESS



 Save

REDACTED FOR CONGRESS



...

[REDACTED] October 10, 2018 · [REDACTED]

There's a new feature (still in beta) that allows one to divide any metrics in GeoInsights by each other. Here's what a map of Instagram for iOS Time Spent/FB Blue Time Spent looks like. Coasts pop out very clearly. (The darkest spots on that map have iOS IG Time-Spent that's 50% of the Blue Time-Spent (across interfaces)).



4

 Like

REDACTED FOR CONGRESS

[REDACTED] has some stuff that may be related: <https://fb.workplace.com/notes/ye-hua/the-state-of-the-states-the-two-many-americas-on-facebook/401294540703903/>

Like · Reply · 1y

[REDACTED] Another piece of work from [REDACTED] specifically about Instagram and coasts.
<https://fb.workplace.com/notes/ye-hua/facebook-in-coastal-us-instagram-teens-and-old-cohorts/462474937919196/>

Like · Reply · 1y



[REDACTED] The earlier image is from [REDACTED] comment on [REDACTED] note here:
<https://fb.workplace.com/notes/spencer-burns/user-diversity-in-the-us-2-demographics/1981172545513273/>

Like · Reply · 1y



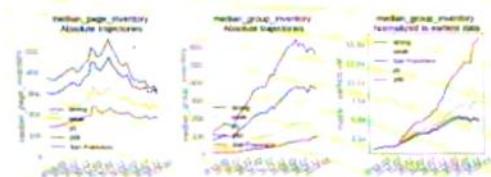
[REDACTED] thoughts i might have on an instagram groups-map you mean?

Like · Reply · 1y

[REDACTED] Right, or some Instagram-equivalent to the idea that local content production isn't correlated with engagement.

Like · Reply · 1y

[REDACTED] Seems like although group engagement is much lower in coastal areas, group inventory is still and only growing in there while being stagnated in the rest of US. So I suspect the longitudinal trajectories of group inventory and group engagement may show a more positive sign for coastal US than a static snapshot.



Like · Reply · 1y



[REDACTED] but if inventory is uncorrelated with engagement, then this wouldn't be the case, no?

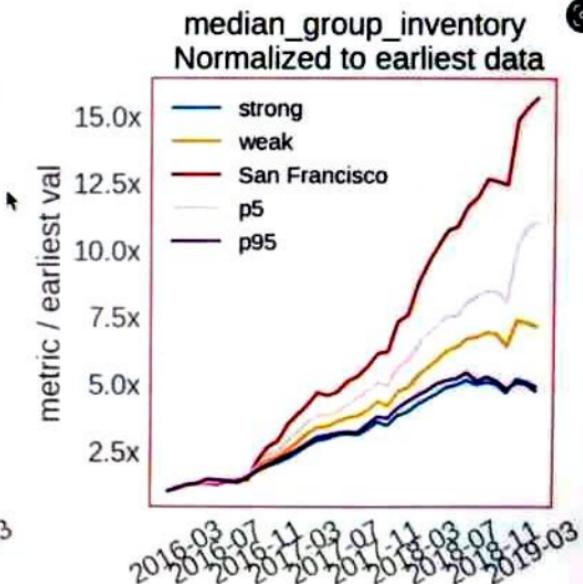
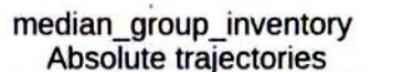
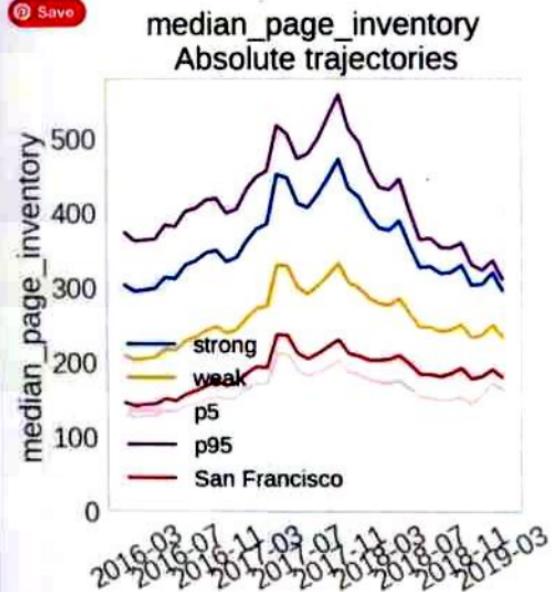
Like · Reply · 1y

[REDACTED] if all group inventory (not just local group inventory) is uncorrelated with group engagement, then yes. I thought this note is focused on local group inventory. Usually inventory and engagement has a positive correlation and then saturates. I am not sure if coastal cities have reach the saturation point.

Like · Reply · 1y · Edited

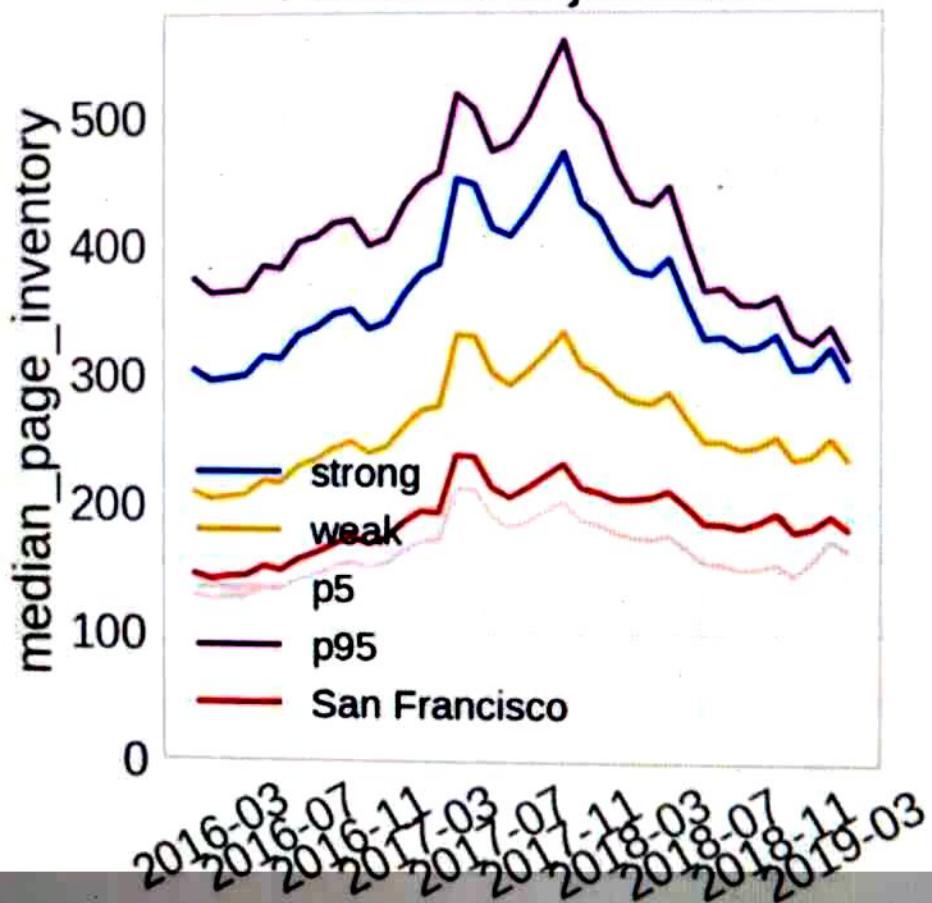
Chats

Save

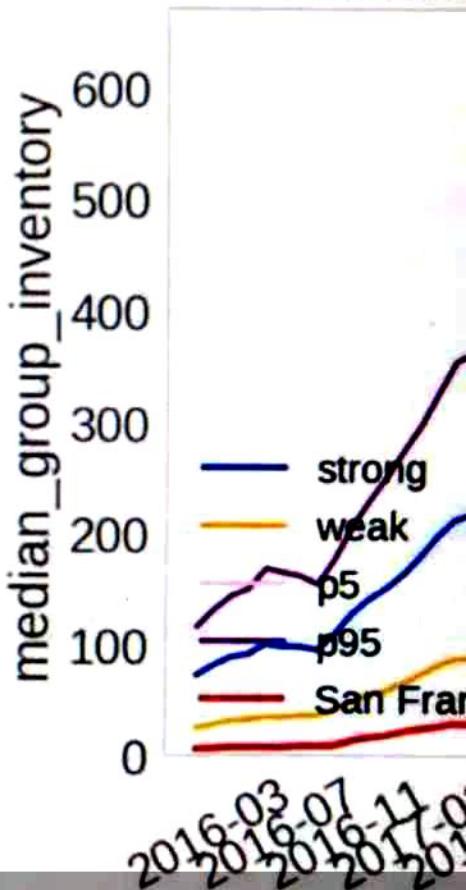


Save

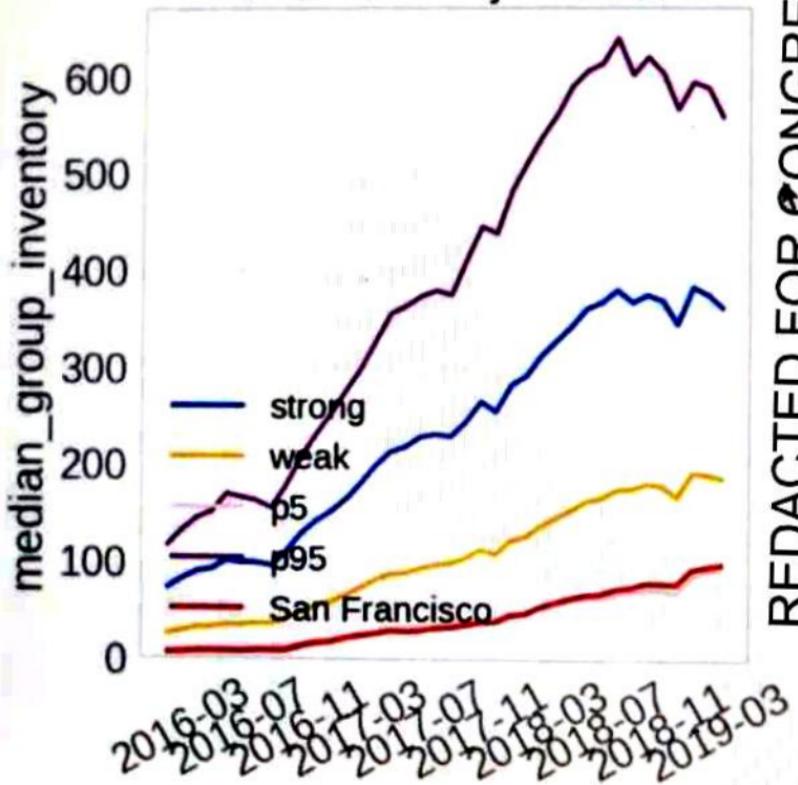
median_page_inventory Absolute trajectories



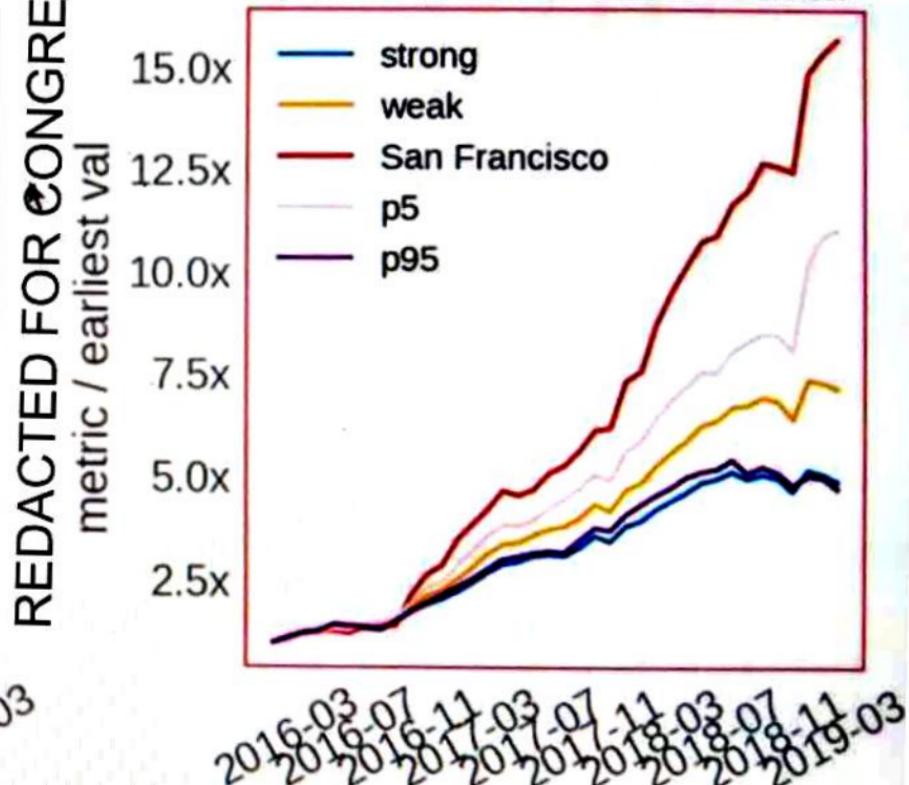
median_group_inventory Absolute trajectories



median_group_inventory
Absolute trajectories



median_group_inventory
Normalized to earliest data



[REDACTED] thoughts i might have on an instagram groups-map you mean?

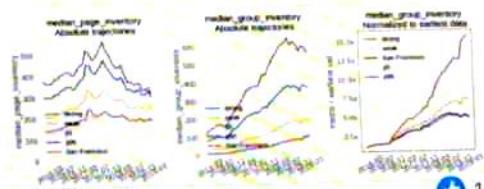
1

Like · Reply · 1y

[REDACTED] right, or some Instagram-equivalent to the idea that local content production isn't correlated with engagement.

Like · Reply · 1y

[REDACTED] Seems like although group engagement is much lower in coastal areas, group inventory is still and only growing in there while being stagnated in the rest of US. So I suspect the longitudinal trajectories of group inventory and group engagement may show a more positive sign for coastal US than a static snapshot.



Like · Reply · 1y

[REDACTED] but if inventory is uncorrelated with engagement, then this wouldn't be the case, no?

Like · Reply · 1y

[REDACTED] if all group inventory (not just local group inventory) is uncorrelated with group engagement, then yes. I thought this note is focused on local group inventory. Usually inventory and engagement has a positive correlation and then saturates. I am not sure if coastal cities have reach the saturation point.

Like · Reply · 1y · Edited

[REDACTED] I thought your graphs are showing local group inventory 😊

Like · Reply · 1y

[REDACTED] oh nah. It is the total group inventory from user_metrics_history.

1

Like · Reply · 1y

Write a reply...

REDACTED FOR CONGRESS

Like · Reply · 1y

1

I thought your graphs are showing local group inventory 😊

Like · Reply · 1y

oh nah. It is the total group inventory from user_metrics_history.

1

Like · Reply · 1y

Write a reply...

Like · Reply · 1y

1

interesting reading!

1

Like · Reply · 1y

One more hypothesis for this: our local group recommendations are the bottleneck, not inventory.

2

Like · Reply · 1y

can you explain this theory? too many local groups, so we recommend bad ones?

Like · Reply · 1y

If more local group inventory isn't causing people to engage more with those groups, then either:

1. Peoples' demand for local groups has been saturated: they're already in enough groups, and don't want to be in more groups.
2. Even if we have more inventory, we don't surface that inventory to people. (ie. the problem is lower down in the funnel.)
3. The incremental local groups are not the types of local groups that people have demand for (this is the hypothesis I suggested in my earlier comment).

I'd love your thoughts,

Like · Reply · 1y

2

really interesting insights! Couple of thoughts

- on user density being negatively correlated with engagement: similar to [REDACTED]'s comment above, but maybe these communities are engaging in real life more easily, meaning that they don't have to engage online as much.
- Given that population density in rural areas is much lower, does having the same threshold for local groups (p75 to centroid of 25km) make sense? These groups would probably be a lot smaller, making it less likely users will engage.

Like · Reply · 1y · Formatted

3

added thought: what were the feature importances of the model you train Chats

1. Peoples' demand for local groups has been saturated: they're already in enough groups, and don't want to be in more groups.
2. Even if we have more inventory, we don't surface that inventory to people. (ie. the problem is lower down in the funnel.)
3. The incremental local groups are not the types of local groups that people have demand for (this is the hypothesis I suggested in my earlier comment).

I'd love your thoughts [REDACTED]



2

Like · Reply · 1y

[REDACTED] really interesting insights! Couple of thoughts

- on user density being negatively correlated with engagement: similar to [REDACTED]'s comment above, but maybe these communities are engaging in real life more easily, meaning that they don't have to engage online as much.
- Given that population density in rural areas is much lower, does having the same threshold for local groups (p75 to centroid of 25km) make sense? These groups would probably be a lot smaller, making it less likely users will engage.



3

Like · Reply · 1y · Formatted

[REDACTED] added thought: what were the feature importances of the model you trained?

Like · Reply · 1y

[REDACTED] I thought I've mentioned it in the post, l28 and friend count are the most important features.

Like · Reply · 1y

[REDACTED] to my second point: one way of seeing if the centroid difference needs to be changed would be by comparing high school class groups in rural areas vs urban. Since school assignment in the U.S is mostly based on geographic proximity, the ratio of the average distance between the two could be a good proxy for how much wider the "local" radius needs to be in different parts of the country. [REDACTED]



2

Like · Reply · 1y · Edited

[REDACTED] that's a cool idea!* I think though to keep this analysis simple, it suffices to count local groups with two radii, and allow the model "choose" according to user density.

(* though, there's quite a bit of problems with school data, ask me how I know)

Like · Reply · 1y

[REDACTED] Write a reply...



1

[REDACTED] Write a comment...



1