Department for Levelling Up,
Housing & Communities

PropTech Engagement Fund Round 1

Final Report

May 2022

GREATER CAMBRIDGE
SHARED PLANNING

Local Planning Authority	Greater Cambridge Shared Planning Service
Date	16/05/22
Theme	**Theme 1:** Plan-making Process: Regulation 18

1.0 Project Aims and Objectives

2.0 Project Summary

3.0 Community Engagement Summary

4.0 Pilot Outcomes

5.0 Development/Implementation

6.0 Procurement

7.0 Conclusions & Policy Reflections

1.0 Project Aims and Objectives

Our project set out to investigate the value of social media insights to inform our consultation topics and local plan focus areas. As part of the intensive consultation program we're carrying out as part of our regulation 18 engagement, we wanted to understand what conversations were happening in and around planning-related topics that weren't necessarily being picked up by the formal consultation process. We call these 'comments in the wild' - topics that are relevant to planning and are incredibly useful but won't be brought up in a formal process for various reasons.

Our project scope was limited to twitter in this instance as our aim was to capture comments that were already made available for public viewing. The assumption is that users turn to twitter to discuss issues in for a public audience and opt into things just as geolocation (you can turn this on and off on will) and relating sentiments to actual events or people (through direct tagging and hashtags). Our aim was to ensure our project did not impose on or listen into conversations that weren't meant to be heard (i.e., such as through private Facebook groups or peoples private WhatsApp messages).

Given this, we set out the objectives of this pilot project to ensure that we could do the following:

- to evaluate the use of social media insights to capture important conversations.
- to cut through to topics people care about that we could do something about as part of the local plan making process.
- to understand where, why, when and how people were tweeting about certain topics.
- to create a use case of best practice into social media insight gathering.

2.0 Project Summary

Our project was designed to use an online, cloud-hosted platform that would allow us to capture public comments and sentiments that aren't necessarily directed to those forming the local plan but could be of use or help make better decisions about planning. The platform had to be capable of the following things:

- the ability to collect data and allow us to filter what was relevant or not;
- the ability to geo-locate the data;
- the ability to visualise in a helpful manner;
- the ability to add sentiment scoring.

Our output would help us develop a gap analysis to understand and assess what is being missed during formal consultation processes. It would ensure our regulation 18 engagement is robust, and that our services truly respond to our public.

We procured The Future Fox, who own the PlaceBuilder platform that specifically looks at sentiment analysis and worked with them over a total of 3.5 months to look at historic data over the past year.

3.0 Community Engagement Summary

Our pilot used historical twitter data collected from UK-based individual users that have some connection to the South Cambridgeshire and Cambridge city accounts over the period of February 2021 – January 2022.

2m tweets were gathered and filtered using a list of location names from areas (parishes, wards, settlements etc) leading to a final 80K. Different selecting criteria were applied subsequently on the 80k tweets to produce the following sub-categories of tweets:

- Keyword: 49,629
- Location: 24,737
- Location with keyword: 11,195
- Location without keyword: 13, 542
- Hashtag: 12,234

These were then used as the basis for our visualisation of sentiment and association over maps. Analysis was calculated using adjusted Natural Language Processing techniques. Associated words in the sentiment over time graph were generated by a deep learning model.

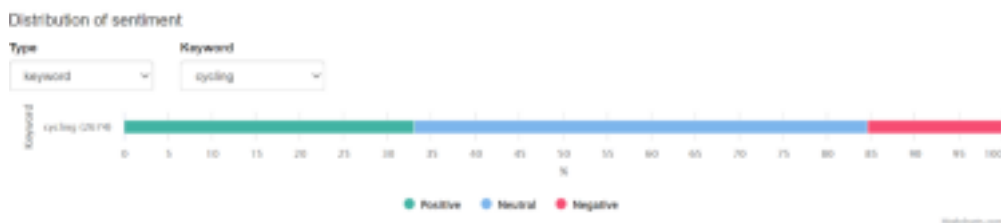This produced three types of visualisations:

1. Where are we tweeting – A geolocated map of tweets:
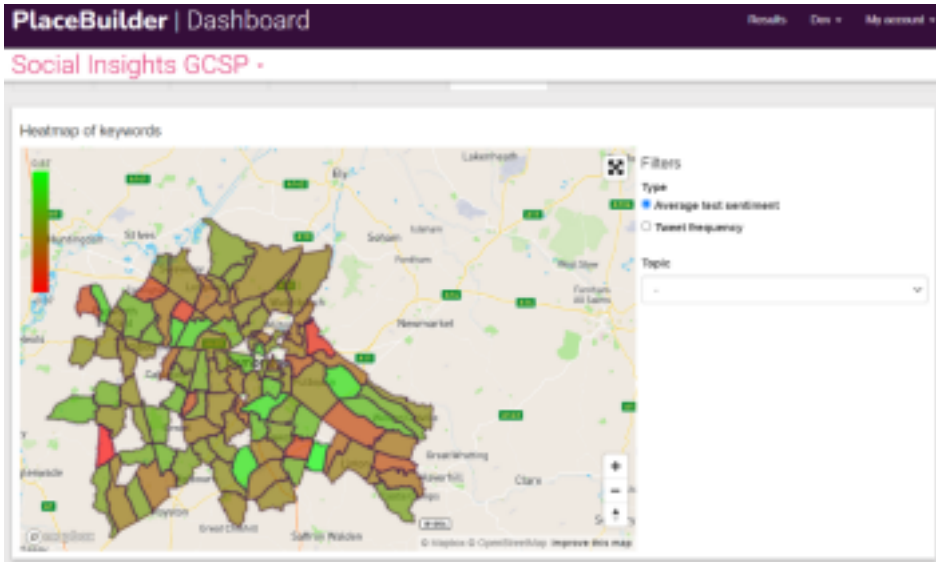
2. When are we tweeting – a visualisation displaying when and what frequency we're tweeting at around these words.

3. How are we tweeting – visualisations on sentiment detected by our algorithms


These visualisations helped us compare various key words to better understand what conversations were happening, what terminology people were using and how they felt (or at least how our algorithm scored their sentiments) around key issues related to planning.

4.0 Pilot Outcomes

The pilot demonstrated the complication in trying to use algorithms and machine learning in figuring out what actually sat behind the data.

For example, looking at the first visualisation which focused on geolocating tweets we have the option of seeing average tweets vs average sentiment of tweets. Using 'congestion' as the first example you can see the following:
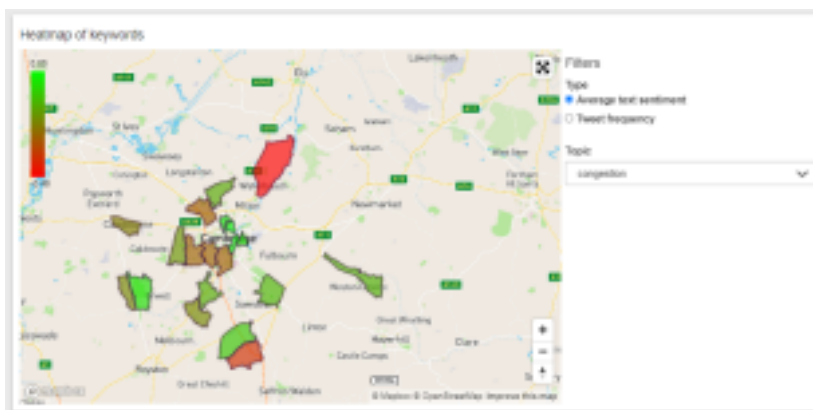
This is average sentiment and below is tweet frequency
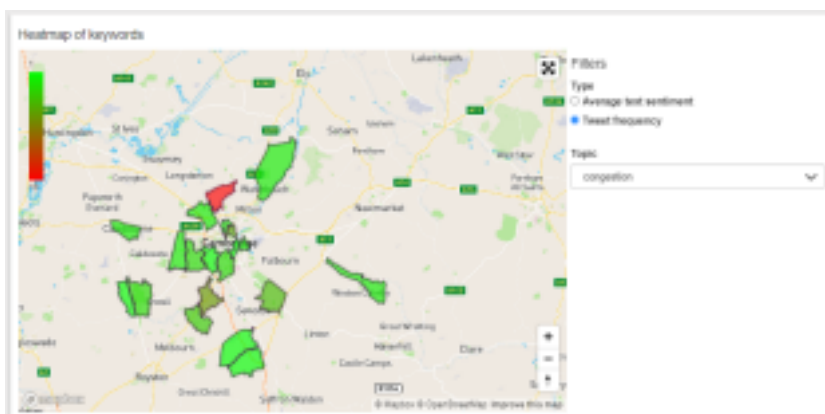
The deep red in the first image shows that tweeters at Waterbeach (the tweets are geolocated by where the tweet was sent from, not necessarily where they were tweeting about or usual place of residence – that would break GDPR rules if we were to do that!). Seem to be more upset about congestion. However the second image shows that there are less people tweeting from there (when hovering over it shows it picks up only two tweets in total). There are more people tweeting from the Histon area, however the first visualisation shows that there is a mixed sentiment towards congestion.

A limitation of this method is that out of the 80K tweets used for this project, only a small percentage can be geolocated as most people have opted out of geolocation.

The second visualisation has gathered the most interest as it demonstrates insights into when certain issues are more relevant.

Using congestion again, versus cycling for example as below:
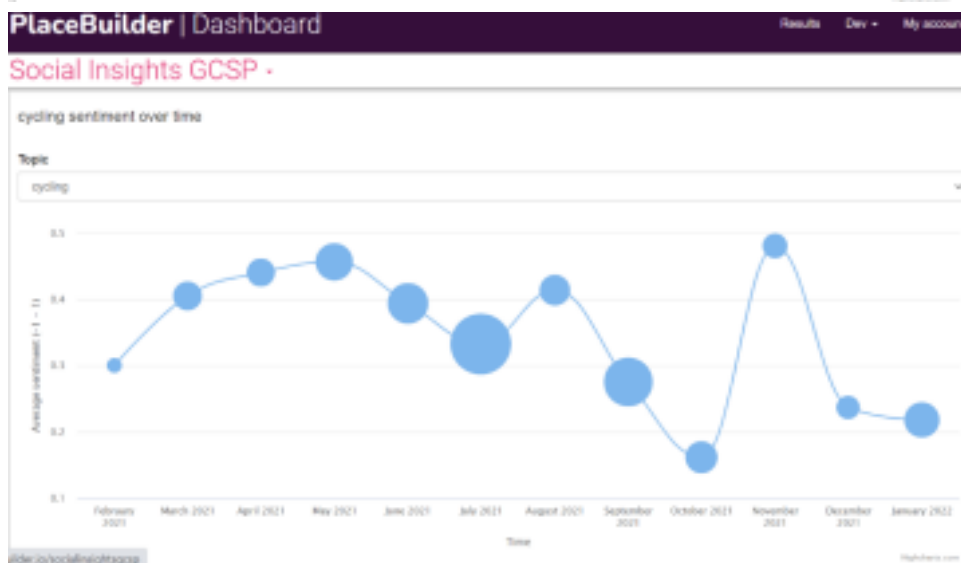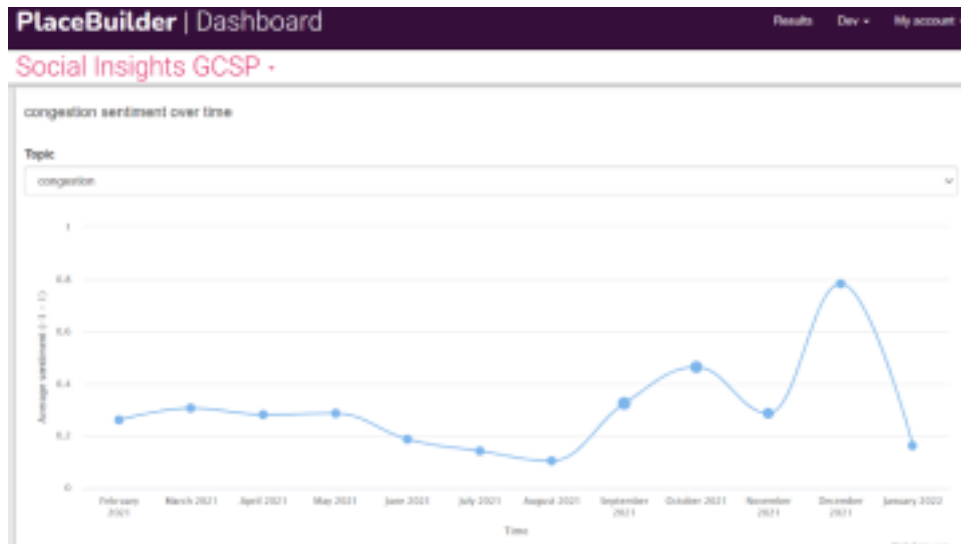
The spike towards the end is where we were consulting on our reg 18 local plan so no surprises there, however over the year more users tweet about cycling (the bubble sizes represent quantum of tweets) and also tend to tweet more positively about it. Also the big bubble for cycling around July 2021 corresponds with the opening of what was a cycle-only bridge during Covid near Mill Road (a heavily congested road that connects to the train station).
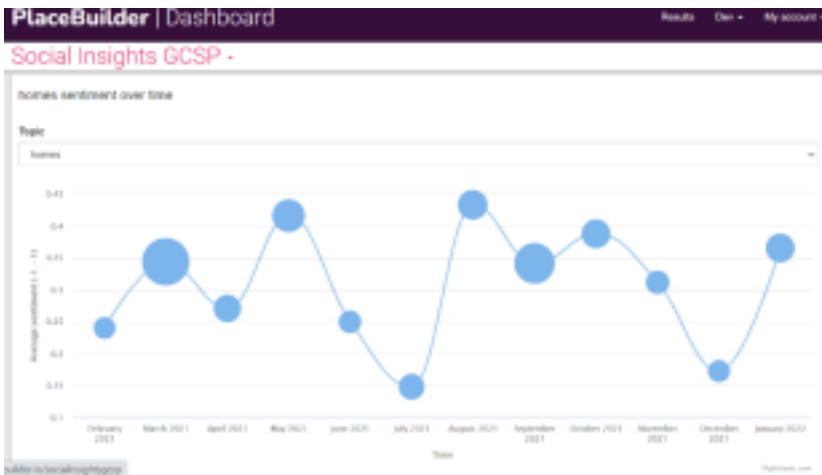
Of course the third visualisation is also vital to cut through the polarity as it demonstrates a more balanced idea of what the algorithm picks up as clearly positive/negative vs just neutral such as below:
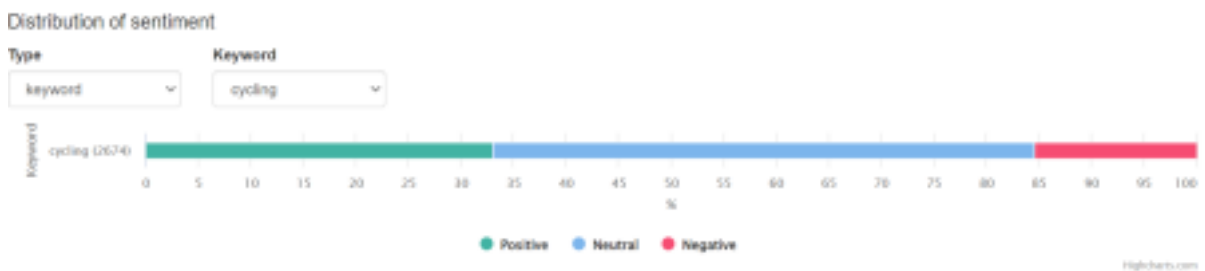
5.0 Development/Implementation

During our development and implementation phases our key findings were:

Finding a starting point amongst all the data is tough - so we start with users. Originally we set out to gather data through known algorithms that generated best matches based on words that were relevant (or we thought were relevant). This poses a problem with spatial data and generally naming data as simple things such as 'water' get confused with places called 'Waterbeach' or the fact that Cambridge UK can get mixed up with Cambridge Massachusetts. To overcome this we took a step back and classified known users and

develop our lists of followers/relevant people. Our starting point ended up being user based rather than keyword based.

Algorithms are only as good as the data you plug through the system and we need people's insights to see where machine learning gets it wrong. In our initial phases we discovered that words like 'graveyard' were being picked up as negative even though the tweets were about how beautiful the old graveyards in Cambridge were. This is why iterative experimentation is important to delivering something that works.

GDPR was also a contentious issue, especially when trying out an exercise that's not really been done for this purpose before. However by tapping into all our resources and connections, as well as sharing learnings across councils we learnt what was acceptable and what wasn't. We found GDPR isn't as much a threat as confirmation bias to our findings. The biggest example when we were looking at tweets around 'water' which we know is an issue, however the data was picking it to be double the amount it actually was because of Waterbeach new town.

6.0 Procurement

Procurement took longer than expected, mainly due to our tight procurement rules. We went through open tender for our platform which meant we had to put resources to developing the brief, the actual procurement process itself and then the evaluation period. We'd dedicated a month to it to be on the safe side which we ended up using in its entirety – something we were only able to do because we had a dedicated Procurement officer to take care of any issues.

We held a supplier briefing which went down well with bidders and also helped the bidders figure out if this was worth their time or not. We had a total of 10 bidders, with a range in quality. Out of these we interviewed the top two and chose the Future Fox who were able to provide us with all the services we needed (and were able to partner with another company for the data collection).

Taking lessons learnt from this we decided any similar project we will try to go via G-Cloud as the contract value isn't worth the time spent on bid writing for bidders or evaluation from our side.

7.0 Conclusions & Policy Reflections

Our conclusion for this project is that though there is use in doing this work, there is still a

long way to go for this kind of project to be incorporated fully into our consultation or engagement strategy. Its experimental nature means there is reluctance from the planners who prefer to focus on what is already in place. This does not mean that its not useful, it has been exceptionally useful for those at management level and our comms team in general.

The pilot as it stands is useful for insights in hindsight and to measure our responses against. We have an idea of how to take this forward if we can get internal consensus and this is still in discussion. Earlier in the project we'd hoped to be able to use this as a way of engaging groups that normally don't engage with us in the formal process, however that needs a lot more than a 6 month project (and technology on its own isn't the answer – we need a more robust and socially-led strategy around it).

The main barriers to take this project forward is the reluctance to rely on social media, even if it were to only form a part of the consultation and engagement strategy. This could be attributed to the fact that no one else is doing this at quite the same scale and there is reluctance to take something this experimental when we're competing with resources. Lack of guidance at national level is also a key issue.