**Summary Introduction**

The government enabled anyone in the UK to set up an official petition, creating an obligation to respond if a sufficient volume of support is created. Petitions normally remain open for 6 months but are rejected if a similar petition is already running.

There are currently[[1]](#footnote-1) 2,805 open petitions (11,889,481 signatures), 1,491 closed petitions (13,075,878 sig), and 51,974 archived petitions (78,957,036 sig); nearly 105million signatures gathered over less than 10 years.

The signatures are different to market research or market activity which is commonly measured by major corporations and marketing companies, particularly to design and execute the promotion of goods and services. Signatures on petitions are the result of a decision made by a particular agent, with a particular location and at a specific time.

There is equally no “code of conduct” or quality regime about how the signatures are obtained but free advice is available[[2]](#footnote-2)[[3]](#footnote-3). A market researcher can maintain a record of people who declined to participate in a survey. But modern online signatures create a high degree of anonymity which has never existed for petitions until very recently.

The government makes the data publicly available, yet there is very little analysis of this large, open data set. Taha Yasseri[[4]](#footnote-4),[[5]](#footnote-5) Bertie Vidgen, Scott A Hale & Helen Z Margetts demonstrated propensity to sign a petition based on a particular topic varies by location.

Stephen D Clark, Michelle A Morris & Nik Lomax used petition data as a way of assessing political leaning by making a comparison against the Brexit referendum, and a classification of voter based on segmentation and geography.

The studies have focused on social and political significance, but opinions and interests also have a commercial interest. Commercial data can be expensive and create a barrier to entry for new businesses and entrepreneurs. Could this open data be exploited for public benefit?

The hypothesis

Can analysis of UK official petition data show interest for different issues in different parts of the country?

To prove this hypothesis, many counter conclusions need to be tested using data science. However, if the hypothesis is disproved, then such a study is likely to demonstrate the key factors that separate successful petitions from unsuccessful petitions – and this would still have an economic and political benefit for the UK public. Such a study would also create a benefit in other countries which have or might consider similar petition systems.

The Task

* Extract data from petition JSON files into a database
* Categorise the headline issues and keywords from each petition.
* Check if there is any correlation between issues and signature volumes.
* Web scrape the number of times the URL to the original petitions have been circulated on the Internet and or social media
* Check if there is a correlation between the number of times a URL shared and the total number of signatures.
* Crosscheck areas for signature counts to see if one area is more likely to sign a petition than another area.
* Crosscheck political dominance in an area to see if this impacts signatures (is an area with an opposition MP more likely to complete a petition signature than an area with a government MP?)
* Crosscheck, size of constituency against signature counts.
* Crosscheck other local demographics such as population density, age or wealth against signature counts.
* Web scrape to see if keywords from each petition are weighted by issues in local areas, then check correlation with petition signatures.
* Investigate if there is a localised element to URL impressions.

Code Sources are available.[[6]](#footnote-6)[[7]](#footnote-7)

Current Assumptions

Keywords from each petition, correlate to topic interest of signatories.

The volume of signatures correlates to the level of interest in an area. We cannot know how many people chose not to sign for reasons other than subject interest, or who were unaware of the petition.

URLs found across the Internet and social media are a proxy of impressions (unless we are unable to check circulation by email or other secure/proprietary sources).

1. <https://petition.parliament.uk/petitions?state=all> [↑](#footnote-ref-1)
2. <https://www.resourcecentre.org.uk/information/organising-a-petition/> [↑](#footnote-ref-2)
3. <https://nationbuilder.com/petition_guide_new> [↑](#footnote-ref-3)
4. <https://epjdatascience.springeropen.com/articles/10.1140/epjds/s13688-017-0116-6> [↑](#footnote-ref-4)
5. <https://link.springer.com/article/10.1007/s11077-020-09395-y> [↑](#footnote-ref-5)
6. <https://bigl.es/getting-petition-data-from-json/> [↑](#footnote-ref-6)
7. <https://petitionmap.unboxedconsulting.com/> [↑](#footnote-ref-7)