Ipsos MORI
Social Research Institute

November 2019

What can we learn about social integration in London from Twitter?

Report for the GLA

Ipsos MORI Social Research Institute

19-013606-01| Version 1 | Confidential | This work was carried out in accordance with the requirements of the international quality standard for Market Research, ISO 20252, and with the Ipsos MORI Terms and Conditions which can be found at http://www.ipsos-mori.com/terms. © Greater London Authority 2019

Contents

19-013606-01| Version 1 | Internal and client use only | This work was carried out in accordance with the requirements of the international quality standard for Market Research, ISO 20252, and with the Ipsos MORI Terms and Conditions which can be found at http://www.ipsos-mori.com/terms. © Greater London Authority 2019

Acknowledgements

This project was commissioned by the Greater London Authority (GLA). Ipsos MORI would like to extend our thanks to Barry Fong, Vivienne Avery, Spencer Thompson and Joe Heywood at the GLA for their insight, advice and feedback throughout the project.

19-013606-01| Version 1 | Internal and client use only | This work was carried out in accordance with the requirements of the international quality standard for Market Research, ISO 20252, and with the Ipsos MORI Terms and Conditions which can be found at http://www.ipsos-mori.com/terms. © Greater London Authority 2019

Executive summary

In response to the Mayor of London's Social Integration Strategy for London, the Greater London Authority (GLA) has begun an ambitious programme to improve its social evidence base and become leaders in measuring issues relating to social integration.

To complement this work, the GLA has set up an 'innovative methods programme' to explore newer types of digital and online data and evidence. The first output of this programme was a scoping study that assessed a range of online data for their relevance and practicality as sources with which to understand social integration. Twitter, a US-based micro-blogging site and public discussion platform, was identified as having potential.

In response to this recommendation, this research sought to measure the sentiment expressed in Tweets sent within and about London across a range of social integration sub-topics. It also sought to assess how the topics and sentiment contained in Tweets differed by demographic and by London Borough.

Finally, the research also aimed to identify a methodology that could be repeated at different points in the future to collect comparable data, so that changes in sentiment or topic frequency could be tracked over time.

This report:

- Provides an analysis of the topics, relevant to social integration in London, that were identified in Twitter data.
- Provides analysis of these topics by age, gender, sentiment and London Borough.
- Presents considerations relating to the representativeness of Twitter users of the wider population of London.
- Provides observations on the challenges and opportunities of using Twitter data to monitor social integration on an ongoing basis.

Takeaways from this report:

- Twitter is a rich source of information relating to social integration in London. Although the analysis did not identify any new areas of social integration which are not accounted for in the GLA's existing framework, it provided anecdotal insight into the impact of social integration; both positive and negative.
- Some areas of social integration are more present in the data than others. Notably, there was a substantial number of topics relating to participation, but far fewer relating to relationships. This may be because these topics are less frequently discussed on Twitter, or an artefact of the methodology.

19-013606-01| Version 1 | Internal and client use only | This work was carried out in accordance with the requirements of the international quality standard for Market Research, ISO 20252, and with the Ipsos MORI Terms and Conditions which can be found at http://www.ipsos-mori.com/terms. © Greater London Authority 2019

- Limited demographic data is made available by Twitter, which limits the potential to conduct robust analysis by factors such as gender and age.
- Findings from Twitter - while of value in their own right - cannot be generalised to the wider population. Only around one in five Londoners (21%) have used or visited Twitter within the last three months and older populations and those of lower social grades are underrepresented.

19-013606-01| Version 1 | Internal and client use only | This work was carried out in accordance with the requirements of the international quality standard for Market Research, ISO 20252, and with the Ipsos MORI Terms and Conditions which can be found at http://www.ipsos-mori.com/terms. © Greater London Authority 2019

Introduction

1.1 Background

The Mayor of London has placed a high priority on improving social integration, equality, diversity and inclusion, and economic fairness across the city. In March 2018, the Mayor of London launched a strategy for social integration 'All of us'[1], with a vision to improve social integration in London through three key areas; relationships, participation and equality.

The strategy also set out the Mayor's framework for understanding and measuring social integration. This included primary quantitative and qualitative data collection to measure social integration across London and aimed to understand specific areas of social integration in more depth.

To address these needs, the GLA has begun an ambitious programme to improve its social evidence base and to become leaders in measuring these issues. The GLA has developed a set of 30 measures[2], that can be used to measure social integration in London. These measures aim to contribute to the GLA's understanding of relationships, participation and equality, by exploring themes such as local civic engagement, feelings of belonging, and unfair treatment.

As part of this research, the GLA has set up an 'innovative methods programme' to explore newer types of digital and online data and evidence to inform this area of work. The first output of this programme was a scoping study[3] that assessed a range of online data for their relevance and practicality of sources with which to understand social integration. One data source recommended by the study was Twitter, a US-based micro-blogging site and public discussion platform.

As a result of this, the GLA wish to explore the extent to which Twitter data provides valuable insight into topics relating to social integration; the topics which are discussed, the sentiment with which topics are discussed, and the extent to which topics differ by demographic characteristics and London Borough. Crucially, the research also aimed to identify an approach to analysing social media data that could be repeated at different points in the future, to collect comparable data, so that changes in sentiment or topic frequency can be tracked.

The specific aims of this project are to develop a deeper understanding of social integration using publicly accessible Twitter data. This will be achieved by:

- capturing Tweets about social integration that are made both in and about these issues in London;

[1] 'All of us: The Mayor's strategy for social integration': https://www.london.gov.uk/what-we-do/communities/all-us-mayors-strategy-social-integration
[2] Social Integration Headline Measures: https://data.london.gov.uk/dataset/social-integration-headline-measures
[3] 'Exploring how to measure social integration using digital and online data https://data.london.gov.uk/dataset/social-integration-digital-online-data-sources

- categorising the topics within these Tweets; and,
- assessing the sentiment expressed across a range of social integration topics.
- when possible, breaking down the topic by demographic information (i.e. age and gender) and geographic location (i.e. London borough)

1.2 Content of this report

This report intends to provide insight into the findings from the data analysis, as well as considerations of the learnings emerging from the methodology. It details:

- The methodology that has been used to export, clean and conduct the topic modelling.
- An overview of the topics, relevant to social integration, that emerged from the topic modelling.
- Analysis of a selection of the topics identified as most relevant to social integration.
- Top-down exploration of additional topics that did not emerge from the topic modelling.
- Considerations about the representativeness of Twitter data.
- Learnings and future considerations.

19-013606-01| Version 1 | Internal and client use only | This work was carried out in accordance with the requirements of the international quality standard for Market Research, ISO 20252, and with the Ipsos MORI Terms and Conditions which can be found at http://www.ipsos-mori.com/terms. © Greater London Authority 2019

Methodology

1.3 Overview

An initial dataset of 400,000 Tweets was collected from Twitter using the social media analytics platform Synthesio[4]. Tweets were collected for the 12-month period between 3 March 2018 and 3 March 2019. To collect the data, a broad Boolean search query[5] was developed, based around London-related terms. The resulting posts were filtered by geo-location metadata and language to ensure that they were London-based and English-language before being exported from Synthesio. To be included, Tweets had to both originate from London (based on geo-tagging and information in users' biographies) and contain a London-related term in the text of the Tweet or Twitter handle. The resulting posts were then cleaned in order to remove irrelevant content. This process created a dataset of Tweets posted in London and about London that were relevant to social integration.

The final cleaned dataset contained circa 50,000 Tweets and was analysed using topic modelling. An initial set of 80 topics were identified. After refining, combination and removal of irrelevant topics, 23 topics that are broadly relevant to social integration were identified.

This chapter provides further detail of the sources, tools, techniques and processes used to collect and analyse the data. Full details of the queries used for the analysis can be found in the appendices.

1.4 Data sources

Ipsos MORI used the social media analytics platform Synthesio to collect Twitter data. Data was collected from Twitter over the 12-month period between 3 March 2018 and 3 March 2019. The entire corpus of Twitter data from this period was searched on the basis of the query that was developed.

Synthesio accessed Twitter data through the PowerTrack API[6]. This allows users to filter Twitter data to topics of interest. An API is an 'Application Programming Interface' which allows different types of software to communicate and exchange data, in this case Synthesio and Twitter. Only public content that was still available at the time of data capture (i.e. had not been deleted from Twitter) was captured.

[4] Synthesio is a social listening platform that sources social data from a range of social media platforms, enriched with robust metadata (such as age, gender and geo-location). Further information can be accessed at: https://www.synthesio.com/.

[5] A Boolean search is a type of search that allows users to combine keywords with modifiers (AND, NOT, OR) to create more specific search results.

[6] Twitter filter https://developer.Twitter.com/en/docs/Tweets/filter-realtime/overview/powertrack-api.html

1.5 Data specification

1.5.1 Initial query development

Relevant Twitter data was collected through a user defined search query developed by Ipsos MORI. A 'query' is a search formula that uses a combination of keywords (which are not case sensitive) and Boolean operators (AND, OR, NOT, NEAR) to isolate information. The query was selectively applied to data within in the Twitter handle. The full query used can be found in the appendices.

To ensure that all Tweets collected referenced a specific London area, the main body of the query was built around the following London related terms:

- Boroughs
- Parliamentary constituencies
- Underground stations
- Overground stations
- National Rail stations
- DLR stations
- Tram stations
- Town centres[7]
- Major parks

There was a large amount of overlap between the terms so, where possible, these terms were simplified and combined. For example, Clapham Common, Clapham High Street, Clapham Junction, Clapham North and Clapham South were all accounted for in the query once, under the key term, 'clapham'.

This approach to building the query – being broad and inclusive instead of homing in on specific social integration-related topics – was taken in order to provide a truer representation of relevant conversations being had in London. It also lessens the possibility of 'force-fitting' any findings based on researchers' pre-conceived expectations of what is present in the Twitter data corpus.

Once the main body of the query had been developed, necessary exclusions that were specific to areas of London were identified. These were used to exclude content that is unrelated to specific areas of London. For example, 'dog' being associated with Barking, or 'bear' being associated with

[7] Town centres were based on those identified in The London Plan 2016: https://www.london.gov.uk/what-we-do/planning/london-plan/current-london-plan/london-plan-2016-pdf

Paddington. These terms were identified by desk research and previewing pilot queries within the Synthesio platform (which did not require data to be downloaded) and undertaking human review.

In addition to the exclusions specific to areas of London, a series of generic exclusions were included in the query. These exclusions were aimed at removing marketing content and news stories, for example terms such as 'win', 'deal' and 'offer'.

1.5.2 Initial query review

Once the initial query had been developed, a random sample of 5,000 Tweets, from 1 January 2019 – 25 February 2019, was exported and manually reviewed. Based on this review, a number of issues were identified.

High proportion of retweets: Over half (55%) of the posts exported were retweets. It is more difficult to infer intention through the action of retweeting a post in comparison to where a user has written a post themselves. Therefore, the decision was made to exclude retweets from the data capture.

High proportion of football related Tweets: The volume of Tweets including terms relating to football clubs was disproportionately high. For example, 14% of the sample of Tweets mentioned Chelsea and 14% mentioned Arsenal. A review of the Tweets mentioning Arsenal confirmed that the vast majority of mentions were related to football. This volume of football related content is problematic as it pushes other content out of the sample and makes other themes more challenging to identify. Therefore, the decision was taken to remove Arsenal, Chelsea, Tottenham and West Ham from the list of key terms included in the query.

Generic London-related terms: Even with place-name specific exclusions, the names of some areas of London (e.g. Embankment, Church End, Bank), were found to return high amounts of content that was unrelated to London. To maintain the relevance to London, the decision was taken to remove such terms from the list of key terms included in the query.

Prolific accounts: Certain categories of Twitter account were found to produce a large volume of content, which was not relevant to the research question. As such Tweets from accounts belonging to London Boroughs, major national news outlets, local news outlets, the metropolitan police, and TfL were excluded from the data collection.

1.5.3 Data collection

Using the Synthesio platform, the dataset was filtered on language (English only), location (London only based on a combination of geo-tagging and information in users' bios) and date the post was made (3 March 2018 – 3 March 2019). A fuller discussion of these exclusions can be found in the 'considerations around representativeness' chapter.

A random sample of 400,000 Tweets from across the 12-month period was exported from Synthesio. Taking a random sample was necessary as the total number of Tweets identified by Synthesio (1,064,812) was outside processing capabilities. In addition to the content of each Tweet, a specific set of metadata was captured and exported. These fields were;

19-013606-01| Version 1 | Internal and client use only | This work was carried out in accordance with the requirements of the international quality standard for Market Research, ISO 20252, and with the Ipsos MORI Terms and Conditions which can be found at http://www.ipsos-mori.com/terms. © Greater London Authority 2019

- Date and time of post
- Sentiment of Tweet
- Age of user (where available)
- Gender of user (where available)

It should be noted that the figures presented throughout the report are not based on individual users as some have multiple posts captured within the dataset. As such, any percentages quoted are based on the number of posts, not the number of individuals posting.

After cleaning the data (as described in the following paragraph), there were 28,120 unique users in the dataset of 50,039 Tweets. It is also important to note that although we are interested in individual Londoners' Tweets, it is not feasible to separate them out completely from people who Tweet on behalf of small organisations, charities or other groups based in London. This means that the dataset also includes content posted on behalf of groups, as well as by individuals.

1.6 Data cleaning and analysis

1.6.1 Additional data cleaning

Following export, machine learning algorithms were used to identify relevant posts. In order to train the algorithm, the research team manually coded a random selection of 2,000 Tweets as either broadly relevant or irrelevant to social integration. The algorithm was run on Neural Networks which recognises patterns and differences in the two groups of Tweets (relevant and irrelevant). Based on this analysis of pre-coded Tweets, the algorithm can make predictions about how to classify new data and modify its categorisation hypothesis through positive and negative reinforcement accordingly. The algorithm achieved 79% accuracy on the test data and so was applied to new dataset of 400,000 Tweets. After data cleaning, the total number of posts in the data set was 50,039. Exclusions mainly related to commercial content which had not previously been detected, transport related content, and some posts which were originally public, but had since been deleted or made private.

The cleaning process was quality assured by manually reviewing a sample of 200 Tweets which had been categorised as 'relevant' by the model, and 200 categorised as 'irrelevant'. Within the automatically judged 'relevant' Tweets, we manually judged that approximately 15% were not relevant (i.e. 85% were relevant). Within the automatically judged 'irrelevant' Tweets, we manually judged that less than 5% were relevant. These levels of accuracy are sufficient and imply that the model is veering towards including Tweets rather than excluding them. This is preferable at this stage in the analysis, as irrelevant topics of conversation could be filtered out at the topic modelling stage.

1.6.2 Topic modelling

Data was analysed using the Ipsos MORI in-house topic modelling platform, built in Python. This used natural language processing techniques to generate a list of terms and phrases that can bring meaning (for example, noun chunks; subject and object in the sentence; terms strongly associated

with other terms). Term similarity was evaluated using a machine-learning algorithm focusing on meaning; words like "good" and "great", for example, were evaluated and classified as similar. Further detail on the topic modelling process can be found in the appendices.

The outcome of this analysis is a topic model – a model reflecting different topics within the data. A single post can be allocated to multiple topics, meaning totals do not always sum to 100%. Initially, 80 topics were generated, which accounted for 80% of the Tweets contained in the cleaned dataset. These topics were qualitatively reviewed by the research team, refined and combined to form 23 topics which were broadly relevant to social integration. Together, these 23 broadly relevant topics, account for 47% of the Tweets contained in the cleaned data set (23,764). The remaining data either could not be categorised (irrelevant or incoherent) or were added to topics which were manually excluded during the refinement topic model process.

1.6.3 Sentiment analysis

Sentiment analysis was run on each topic. A sentiment is the indicator that determines whether a post contains positive, negative, or neutral language. This is assigned automatically in this research by a patterned algorithm[8] which analyses the number of positive or negative words and emojis in a post, based on a predetermined list (e.g. good, great, amazing etc). Each part of a post is analysed for sentiment meaning a single post can be classified multiple times. For example, a post which contains a sentence complaining about inaccessible transport, followed by a sentence which talks about great customer service from TfL staff, may be categorised under both the positive and the negative sentiment category. However, as Twitter posts have a 280-character limit and the average length of a Tweet is 33 characters, the impact of this is small as Tweets are less likely to contain multiple sentiments.

It is important to note that the length of many of the posts and the wide variety of content and language used means that the value of sentiment analysis for analysing the data is limited. This type of analysis does not always capture the nuances of the English language. For example, the algorithms do not consistently identify sarcasm, or swear words being used to express positive sentiment. Certain thresholds must also be met for posts, or sections of posts, to be categorised under a sentiment meaning not everything will be captured. As will be noted throughout the report, for all topics, most Tweets were classified as neutral.

1.6.4 Analysis by age and gender

Analysis by both age and gender was applied to each topic. This analysis utilised the age and gender metadata that Twitter makes available (where users have provided this information in the profile). The conclusions that can be drawn from this data are limited however because it is only available for a relatively small subset of Twitter users; age was available for 45% of Tweets within the cleaned dataset while gender was available for 20% of users in the cleaned dataset. Furthermore, age is only available by three very broad age bands (under 18 years, 18-29 years, 30 years or older). This limits the insight

[8] Further information available at: https://www.clips.uantwerpen.be/pages/pattern-en#sentiment

that analysis by age can generate, particularly for older age groups. Due to these considerations, analysis by age and gender should be treated as indicative only.

The following tables show the breakdown of age and gender across the 23,764 Tweets contained in the 23 topics relevant to social integration. A relatively high proportion of Tweets (16%) originated from users aged below 30. Among the small proportion of Tweets for which the user's gender is known, more were from male users (12%) than female users (7%).

Age	Count	Percentage of Tweets (all)	Percentage of Tweets (known)
Under 18 years	1,215	5%	11%
18-29 years	2,550	11%	24%
30 years or older	6,887	29%	65%
Unknown	13,112	55%	N/A
Total	23,764	100%	100%

Base: All Tweets containing content related to one or more of the 23 topics relevant to social integration (23,764).

Gender	Count	Percentage of Tweets (all)	Percentage of Tweets (known)
Female	1,714	7%	37%
Male	2,968	12%	63%
Unknown	19,082	80%	N/A
Total	23,764	100%	100%

Base: All Tweets containing content related to one or more of the 23 topics relevant to social integration (23,764).

The following table shows a demographic breakdown of the 4,184 Tweets (within the 23,764 relevant Tweets) for which both age and gender of the user was known. This breakdown demonstrates that males aged 18-29 are likely to be over-represented in the dataset: while estimates suggest they make up approximately 9% of London's population[9], they contribute 15% of relevant Tweets. In contrast, women aged 30 year or older appear to be under-represented: estimates suggest they make up approximately 30% of London's population, but they only contribute 24% of the relevant Tweets.

[9] Estimates based on the 2016-based GLA population and household Central trend projections for 2019: https://data.london.gov.uk/dataset/projections

Age	Gender: Female	Gender: Male	Total
Under 18 years	4%	5%	10%
18-29 years	9%	15%	23%
30 years or older	24%	43%	67%
Total	37%	63%	100%

Base: All Tweets containing content related to one or more of the 23 topics relevant to social integration, where both age and gender or Twitter user is known (4,184).

1.6.5 Top-down analysis

The bottom-up topic modelling succeeded in identifying many of the key elements of social integration within the cleaned data corpus. Notably, there were substantial quantities of topics relating to participation. However, it was notable that some topics of interest were missing from the topic model; particularly those relating to relationships.

To validate the topic model, and check whether this data is present in the data corpus, we conducted top-down analysis on key topics relating to social integration which were not present in the topic model:

- Helping neighbours
- School activities and events
- Conversations with friends
- Local area change and affordability (this might include shops closing and new housing developments)

To conduct this analysis, a series of Boolean search queries were written (see appendices for full queries), containing key terms relating to the topics of interest. These queries were then run on the cleaned data corpus of 50,039 Tweets and refined further as necessary.

1.6.6 Qualitative analysis

Qualitative review and analysis were applied throughout the research, particularly for refining the topic modelling, by experienced Ipsos MORI researchers. Qualitative review was essential to quality assure the topic model, further unpick the themes identified by the topic model, and interpret the meaning and nuances within the data.

During the topic model refinement process, researchers reviewed a random sub-set of Tweets which had been allocated to each topic. The primary purpose of the qualitative review was to assess the extent to which the Tweets assigned to each topic formed a cohesive narrative. Tweets were coded as either relevant, or irrelevant to the topics. Where a notable proportion of Tweets within each topic were judged to be irrelevant to the topic, the term-similarity threshold of the topic was increased to

19-013606-01| Version 1 | Internal and client use only | This work was carried out in accordance with the requirements of the international quality standard for Market Research, ISO 20252, and with the Ipsos MORI Terms and Conditions which can be found at http://www.ipsos-mori.com/terms. © Greater London Authority 2019

heighten relevance (by reducing variation within the Tweets included in the topic). In some cases, the Tweets included in separate topics were noted to be very similar and, in these cases, topics were combined.

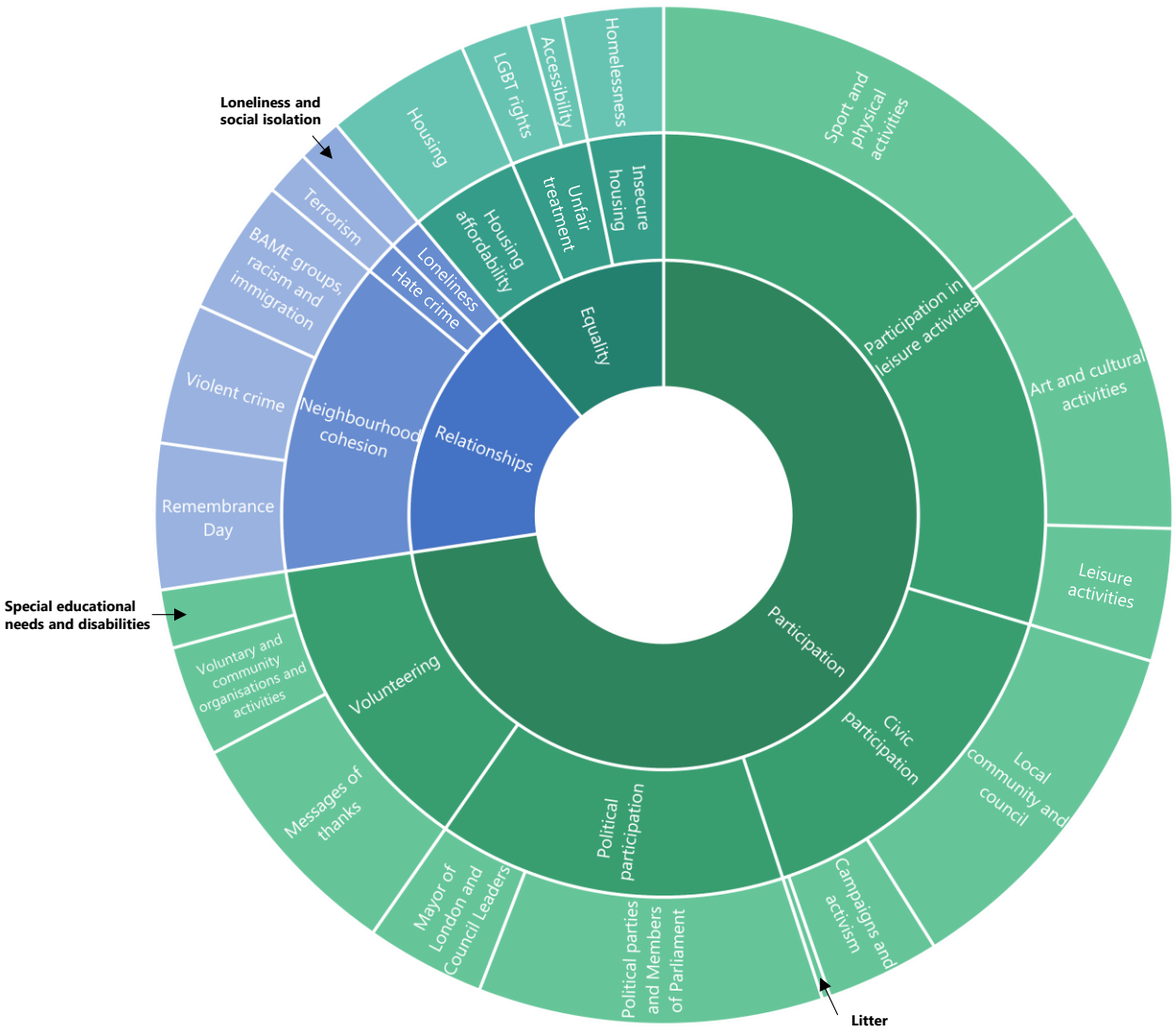
Following refinement of the topic model, thematic qualitative analysis was conducted on the Tweets within each topic. A random sample of the Tweets that had been allocated to each topic were reviewed in full, to identify interesting aspects of the data that could form the basis of sub-topics. Tweets were reviewed until it appeared saturation had been reached; no new sub-topics were emerging. The content included in some topics was relatively uniform; resulting in few sub-topics, while larger topics tended to contain Tweets relating to a range of sub-topics. This kind of qualitative analysis, by its interpretive nature, involves a degree of subjectivity. Ipsos MORI researchers interpreted the data, framing their analysis with reference to the research objectives and background to the research. This involved making judgements about which topics to include.

This qualitative approach is also reflected in the way the data is reported. Although numerical data is provided on the attitudes and characteristics of Twitter users where appropriate, for the most part a qualitative approach to findings has been taken.

19-013606-01| Version 1 | Internal and client use only | This work was carried out in accordance with the requirements of the international quality standard for Market Research, ISO 20252, and with the Ipsos MORI Terms and Conditions which can be found at http://www.ipsos-mori.com/terms. © Greater London Authority 2019

Describing the conversation

This chapter outlines the key findings from the analysis of Twitter posts captured as part of the query, as described by the topic model. The topic model identified 23 topics, which were judged by the research team as being broadly relevant to social integration (as defined by the GLA's social integration measures). The following topic wheel shows 20 of these topics (outermost layer of the topic wheel), and gives an indication of the relative size of each topic and broadly how they map to the social integration measures (middle layer of the topics wheel) within the three main domains (innermost layer of the topic wheel) outlined by the GLA. Although in most cases the identified topics correspond to the GLA's social integration measures, as they were generated using a bottom-up approach-without reference to the measures-there are some instances where topics' correspondence to the measures is imperfect. A further three topics (appreciation of public spaces, Brexit, and mental health and wellbeing), which were identified by the topic model but do not correspond with the GLA's social integration measures, are shown in the table on the next page.



For the fully cleaned data set, 47% of posts were categorised into topics described by the topic model. The table below shows these topics, including those that relate to the three main domains identified by the GLA (topics included in the topic wheel above), as well as three further topics identified by the model that relate to social integration but do not fit into the three domains (appreciation of public spaces, Brexit, and mental health and wellbeing). Please note that the percentages reported in the following table are based on all posts that were categorised, rather than all posts in the final cleaned data set. Please also note that, as Tweets could be multi-coded, many will appear in more than one topic. Percentages should be treated as broadly indicative of volume only. The topics in the rows that are shaded have been identified by the GLA as particularly relevant to social integration and are discussed in more detail in the following chapter.

Topics relating to the Relationships domain

Topic name	Theme	Description	Proportion
Black, Asian and minority ethnic groups, racism and immigration	Neighbourhood cohesion	Mentions of Black, Asian and minority ethnic groups (both positive and negative), racism, hate, antisemitism, cleansing and immigration.	5%
Remembrance Day	Neighbourhood cohesion	Mentions of Remembrance Day	5%
Violent crime	Neighbourhood cohesion	Mentions of knife crime	5%
Loneliness and social isolation	Loneliness	Mentions of organisations working to tackle social isolation and loneliness, including those aimed at the elderly, the young and disabled people	2%
Terrorism	Hate crime	Mentions of terror attacks and violent crime, including specific attacks (e.g. Westminster), extremism and racism	2%

19-013606-01| Version 1 | Internal and client use only | This work was carried out in accordance with the requirements of the international quality standard for Market Research, ISO 20252, and with the Ipsos MORI Terms and Conditions which can be found at http://www.ipsos-mori.com/terms. © Greater London Authority 2019

Topics relating to the Participation domain

Topic name	Theme	Description	Proportion
Sport and physical activities	Participation in leisure activities	Mentions of participation in sport and physical activities including cycling, walking, running, swimming and yoga.	18%
Local community and council	Civic participation	Mentions of plans and proposals for the local area, residents and the community	14%
Political parties and Members of Parliament	Political participation	Mentions of key political parties and party members, elections and voting	13%
Art and cultural activities	Participation in leisure activities	Mentions of local activities, including art, poetry, museums, and music events. Mentions of diversity and multicultural events	12%
Messages of thanks	Volunteering	Messages thanking others for support, charity donations and volunteering	9%
Mayor of London and Council Leaders	Political participation	Mentions of the Mayor of London and Council Leaders	5%
Leisure activities	Participation in leisure activities	Mentions of day-time/family/weekend activities	5%
Voluntary and community organisations and activities	Volunteering	Mentions of voluntary and community organisations and activities, including young people, those with disabilities and those with mental health problems	4%
Campaigns and activism	Civic participation	Mentions of campaigns and petitions to support a wide range of issues, including council services	4%
Special educational needs and disabilities	Volunteering	Mentions of raising awareness of SEND and services available. Includes mentions of children, young people, the elderly and carers	2%
Litter	Civic participation	Mentions of litter-picking and discouragement of littering	*

19-013606-01| Version 1 | Internal and client use only | This work was carried out in accordance with the requirements of the international quality standard for Market Research, ISO 20252, and with the Ipsos MORI Terms and Conditions which can be found at http://www.ipsos-mori.com/terms. © Greater London Authority 2019

Topics relating to the Equality domain

Topic name	Theme	Description	Proportion
Housing	Housing affordability	Mentions of housing shortages and affordability, new builds, landlords vs. renters' rights	6%
Homelessness	Insecure housing	Mentions of homelessness, people harassing the homeless, stealing their clothes, tents and sleeping bags	4%
LGBT rights	Unfair treatment	Celebration of LGBT rights and the LGBT community	3%
Accessibility	Unfair treatment	Mentions of the limitations and problems disabled people face accessing transport and public facilities	1%

Topics relating to social integration outside of the three main domains

Topic name	Theme	Description	Proportion
Appreciation of public spaces	N/A	Mentions of enjoying being outside in public places	9%
Brexit	N/A	Mentions of the EU, UK government and concerns surrounding Brexit	9%
Mental health and wellbeing	N/A	Promotion of mental health and wellbeing initiatives	9%

Note: * indicates less than 1%.

The frequencies, even among the largest topics, are relatively small. This demonstrates that no single topic dominates online conversations about social integration topics. This also reflects the diversity of social media conversations – there is a huge variety in terms of types of issues and language used.

19-013606-01| Version 1 | Internal and client use only | This work was carried out in accordance with the requirements of the international quality standard for Market Research, ISO 20252, and with the Ipsos MORI Terms and Conditions which can be found at http://www.ipsos-mori.com/terms. © Greater London Authority 2019

Key topics of conversation

This chapter provides narrative description of each of the topics which were identified by the GLA as particularly of interest:

- Black, Asian and minority ethnic groups, racism and immigration
- Loneliness and social isolation
- Local community and council
- Voluntary and community organisations and activities
- Art and cultural activities
- Messages of thanks
- Homelessness
- Housing
- LGBT rights
- Accessibility

Where tweets are shown, these have been redacted and altered as necessary to retain anonymity.

19-013606-01| Version 1 | Internal and client use only | This work was carried out in accordance with the requirements of the international quality standard for Market Research, ISO 20252, and with the Ipsos MORI Terms and Conditions which can be found at http://www.ipsos-mori.com/terms. © Greater London Authority 2019

1.7 Relationships

1.7.1 Black, Asian and minority ethnic groups, racism and immigration

A relatively small proportion (5%) of the relevant Tweets mentioned ethnic minorities, racism and immigration. The word cloud below illustrates the words used in Tweets relating to these topics. Within this topic, the term 'colour_name' refers to mentions of 'black' or 'white'.

Most Tweets around this topic were identified as neutral (79%) or positive (19%), although a small proportion (1%) did display a negative sentiment. This is an example of where sentiment analysis has struggled to accurately categorise tweets. For example, mentions of 'cleanse', in relation to ethnic cleansing are negative, but were not identified as such by the algorithm (as the word 'cleanse' is not intrinsically negative). Most Tweets in this topic promoted events or courses that aimed to educate people on the history of minority ethnic groups or on specific issues faced by minorities. Some discussed racism in politics or immigration policies, while others condemned racism they had seen or heard. A proportion of Twitter users also made it clear that they didn't welcome or embrace London's ethnic diversity, with some talking about the 'ethnic cleansing' of White people.

Celebrating diversity

Several Tweets expressed general appreciation towards diversity in London. Some Tweets discussed the representation of ethnic minorities in certain areas such as politics. For example, there were Tweets congratulating organisations for having diverse leaderships boards, or councils for having representatives of different ethnicities. There were also ideas for how to showcase achievements of

people from ethnic minority backgrounds, in order to raise BAME children's aspirations. However, some people also called for more diverse representation.

"It's the people that make a neighbourhood - and London is made of immigrant stories."

"There's huge diversity in [borough]. Walk the streets and see residents getting on. A white guy fell in road & migrants rushed to help. Brits aren't Racist. Most people aren't Racist. Beware the Haters."

Condemning racism

Some Tweets commented on racist language or behaviour. These included examples of racism people had seen or experienced, with reactions of disgust, anger or disappointment. Some of these Tweets also called for action in response to these incidents, such as boycotting organisations associated with the incidents.

"Another day another racist in [supermarket] in [borough] telling a black employee that he doesn't belong in England"

"Could you please boycott the newsagents in [neighbourhood]. They racially abused a BAME customer this morning"

Negativity towards immigration

Tweets that expressed negativity towards immigration and ethnic minorities often mentioned experiences of 'ethnic cleansing' of White people. Some people blamed specific individuals or groups, such as the political parties, for ethnic cleansing in particular areas.

The BBC was also criticised for being biased in reporting issues relating to minority ethnic groups and not reporting on the cleansing of White people in London.

"When is BBC going to report on the racist cleansing of white people from [political party] London?"

Similarly, some users conveyed annoyance at positive discrimination toward minorities or hypocrisy in judging British people and immigrants differently. For example, several users claimed that immigrants are being prioritised before British nationals for housing allocation.

"Hate speech is only applicable if you're a so-called minority, but where there are lots of immigrants and they are insulting to the British it's not hate crime."

There were also more general comments about immigration policy, including discussions about political parties' or local politicians' stance on immigration.

"Will [borough] and particularly [parliamentary constituency] reduce uncontrolled immigration, solve the housing crisis, the NHS deficit, and fewer school places."

Education

Several Tweets mentioned education relating to BAME history, immigration or specific issues. The Windrush generation was mentioned, for example when events had been organised about the topic.

"We had a great time today at [primary school] speaking in assembly about the Windrush Generation and finding out about the experiences of Caribbean elders in [borough]"

Some users also commented on the lack of education about BAME history in schools, or the lack of discussion in the news about issues experienced by particular minority groups.

"I find it strange that schools don't teach black British history. It continues the narrative that there is no racism in Britain and that there was no civil rights movement"

Comments relating to ethnicity, racism and immigration were most prominent in Barking and Dagenham (11%) and Brent (13%), where they comprised over 10% of the conversation around social integration. Among Twitter users where gender was known, the topic was similarly likely to be mentioned by male users (5% of males mentioned this topic), as by female users (4%)[10]. Broadly similar proportions of those aged below 18 (5%), 18-29 (3%) and 30 and over (6%) mentioned the topic.

1.7.2 Loneliness and social isolation

A small proportion of the relevant Tweets (2%) addressed the topics of loneliness and social isolation. Around one in three (32%) of these Tweets displayed a positive sentiment, while just 1% had a negative sentiment. The word cloud below illustrates the words used in Tweets relating to these topics.

[10] Analysis by gender and age should be treated as indicative only due to the very low availability of gender and age information available in the dataset.

19-013606-01| Version 1 | Internal and client use only | This work was carried out in accordance with the requirements of the international quality standard for Market Research, ISO 20252, and with the Ipsos MORI Terms and Conditions which can be found at http://www.ipsos-mori.com/terms. © Greater London Authority 2019

This topic covered efforts made by both organisations and individuals to reduce social isolation and loneliness. The negative impact social isolation has on mental health was raised in a selection of Tweets under this topic. There were also mentions of events being held by charitable organisations such as Mind, British Red Cross and Age Concern UK to alleviate social isolation and loneliness. Multiple Tweets raised awareness of smaller community projects and activities. On occasion, Tweets made specific reference to the problem of loneliness and social isolation among the elderly and disabled people.

Organisations and groups tackling loneliness and social isolation

Most Tweets relating to loneliness and social isolation came from groups rather than individuals, both large national organisations and smaller community groups. These groups were Tweeting about a wide range of projects or activities that aimed to help vulnerable or isolated people. Some groups did not refer to loneliness specifically, but organised activities that promoted socialising and meeting new people, such as groups for new parents.

"Whether it's sunny, or it's raining, people still feel lonely. Please visit [website] to find out more how we're tackling loneliness and isolation in our community"

"It's possible to be surrounded by people but feel lonely. It's a massive, busy city. We're out & about in [neighbourhood], asking Londoners how they're feeling"

Support for elderly residents

Many of the projects and activities users Tweeted about were aimed at supporting elderly people in their communities. The support offered ranged from volunteers organising one off events for elderly residents - such as Christmas dinners - to ongoing programmes, such as intergenerational mentoring schemes.

"How lovely is this. Local people in [district]help to cook dinner for an elderly neighbour. Thanks [user handle] for being a giving human."

"Pairing young people with an elderly person who has similar interests can reduce loneliness, give people purpose and develop key skills in [borough]"

Training and educational opportunities

Some organisations Tweeted about training and educational courses they were running. For example, there were courses that aimed to raise the public's awareness of vulnerable groups and the specific issues they face, or tackle misconceptions about groups such as those with dementia. Other courses were designed to support specific groups and develop skills, such as a course to help elderly people stay in control of their finances or a stress and anxiety management course.

"At [borough council] dementia & social isolation training today. One of our 'befriendees' is telling us about how his friend has improved life for him and his partner who were previously very isolated."

Raising funds

Many Tweets asked people to donate money to support charities or projects that reduced social isolation.

"We're fundraising to support isolated older people in [borough]. We will help older people enjoy later life & provide support and fun community events to promote good mental health."

Tweets tagged as belonging to the boroughs of Waltham Forest (4%) and Tower Hamlets (3%) were particularly likely to mention topics relating to loneliness and social isolation, though it should be noted that, compared to other topics, the number of mentions of loneliness and social isolation showed relatively little variation across boroughs. Tweets relating to loneliness and social isolation were posted more frequently by females (accounting for 2% of Tweets by females) compared with the proportion of males who mentioned this topic (1%)[11]. However, the topic was equally likely to be mentioned across age groups (mentions of the topic accounted for 2% of Tweets from those under 18, those aged 19-29, and those aged 30 or over).

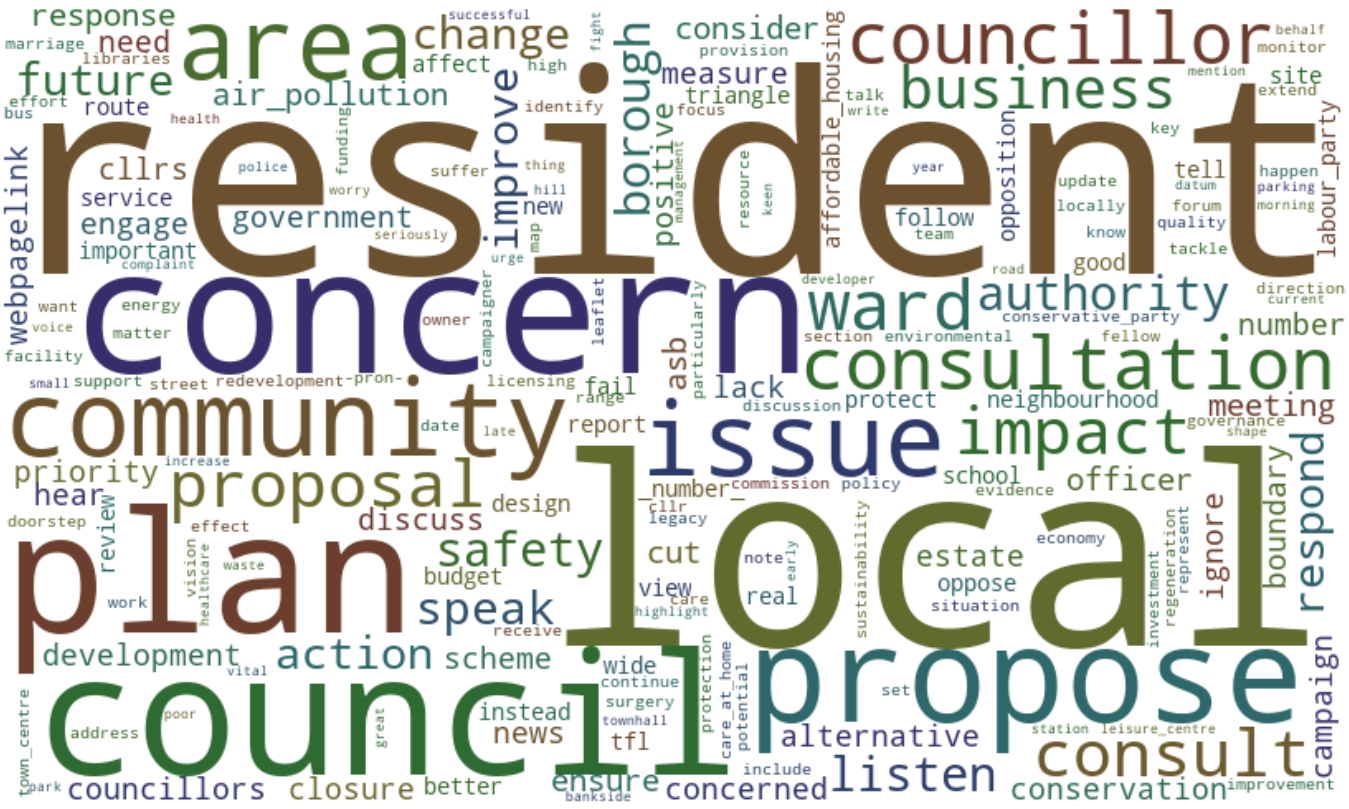
[11] Analysis by gender and age should be treated as indicative only due to the very low availability of gender and age information available in the dataset.

19-013606-01| Version 1 | Internal and client use only | This work was carried out in accordance with the requirements of the international quality standard for Market Research, ISO 20252, and with the Ipsos MORI Terms and Conditions which can be found at http://www.ipsos-mori.com/terms. © Greater London Authority 2019

1.8 Participation

1.8.1 Local community and council

Fourteen per cent of the Tweets addressed the topics of local community and council. One quarter of these Tweets (24%) displayed a positive sentiment, while three quarters (76%) displayed a neutral sentiment. The word cloud below illustrates the words used in Tweets relating to local community and council:



A number of Tweets mention topics that demonstrate local residents' engagement with their local communities and local government. A wide range of issues relating to local government are covered in these Tweets. Some expressed positivity and gratitude towards their council for making their borough a pleasant place to live. Others contained a more negative sentiment and complaints were varied.

Encouraging civic participation

A subset of Tweets in this topic aimed to raise awareness and encourage participation in a wide range of public consultations. The consultations mentioned include those on changes to bus routes, cuts to special needs funding in schools, changes to urgent care services, and speed reduction measures. A small number of Tweets criticised consultation processes or the way that consultation data was used.

"Consultation in [borough] last night on @TfL's local plans for walking and cycling"

Several Tweets aimed to raise awareness and encourage participation in residents' associations; giving information on upcoming meetings and the topics that would be under consideration.

"If you live in [neighbourhood], join your local residents' association to improve air quality in [borough]"

Council funding

Other Tweets concerned decisions councils had made about funding or were petitioning for more funding. This included funding for special needs in schools, police, education in general and to redevelop tube stations and leisure centres.

"It was good to see so much support for our campaign for safer streets and more police funding!"

Several Tweets also drew attention to the need for funding for additional affordable housing for residents, and related planning applications.

"It's taken a long time, but what a brilliant achievement! New community-built council homes for local residents in [neighbourhood]. Genuinely affordable housing. Great work everyone!"

Local air quality

Several Tweets expressed concern about the air quality in their local area. Others offered potential solutions to tackling the problem or encouraged others to becoming involved in improving air quality.

"We will be at community centre tonight for [borough council] residents' engagement on clean air. We continue to work for better air quality, as part of our commitment to the environment"

Mentions of local community and council were particularly frequent in the outer London Boroughs of Hillingdon (where 20% of relevant Tweets referred to these topics), Bexley (19%), Merton (19%), Redbridge (18%) and Sutton (19%).

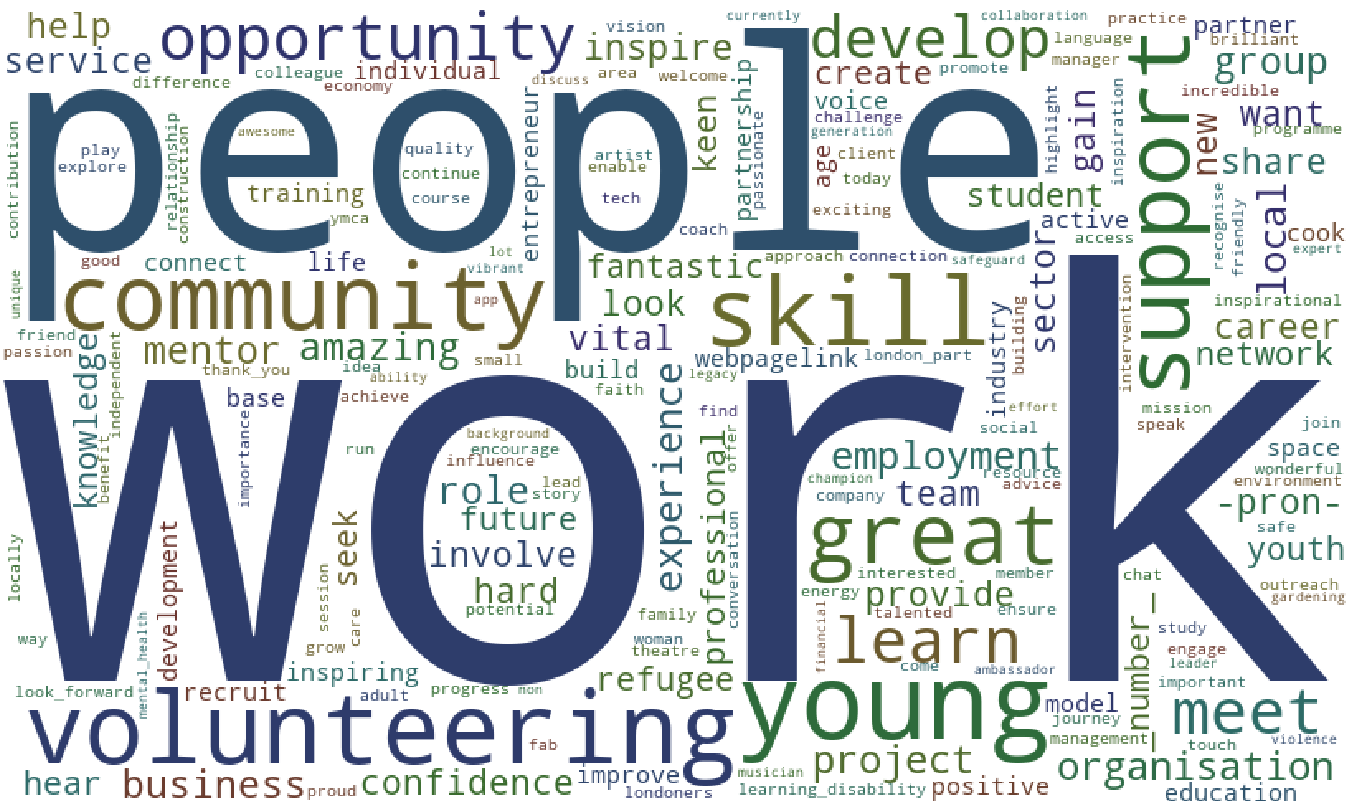
Turning to look at this topic by gender, males and females were similarly likely to engage with local community and council issues on Twitter[12] (14% of males and 12% of females mentioned this topic). Analysing the topic by age, it seems that those aged under 18 (for whom mentions relating to community and council account for 11% of their relevant Tweets), those aged 18-29 (14%) and those aged 30 or over (10%) are similarly likely to mention this topic.

[12] Analysis by gender and age should be treated as indicative only due to the very low availability of gender and age information available in the dataset.

19-013606-01| Version 1 | Internal and client use only | This work was carried out in accordance with the requirements of the international quality standard for Market Research, ISO 20252, and with the Ipsos MORI Terms and Conditions which can be found at http://www.ipsos-mori.com/terms. © Greater London Authority 2019

1.8.2 Voluntary and community organisations and activities

Four per cent of the Tweets addressed the topics of voluntary and community organisations and activities. One third (33%) of these Tweets displayed a positive sentiment, while fewer than 1% had a negative sentiment. The word cloud below illustrates the words used in Tweets relating to voluntary and community organisations and activities:



Tweets within this topic covered a range of themes relating to voluntary and community organisations.

Volunteering opportunities

Some Tweets originated from voluntary and community organisations themselves and promoted opportunities for volunteering with them. These organisations included those supporting children and young people, working with people with learning disabilities and working with homeless people among others.

"Looking for work or a volunteering opportunity? Visit us in [shopping centre] today to find out more about supporting people into work #LearningDisabilityWeek"

"We are at the volunteering [borough] Fair with other local agencies, celebrating our wonderful volunteers whilst encouraging visitors to join our highly recognised volunteer programme"

Similarly, other Tweets within this category drew attention to, and encouraged participation in, specific community activities such as community gardening and craft classes.

"A father and child were playing near the garden and we invited them to sow some seeds. The first residents to work in our new community greenhouse."

Benefits of volunteering

Other Tweets promoted the benefits of volunteering. For example, some detailed how volunteering experience had helped individuals to gain paid employment, while others mentioned how volunteering can help individuals feel more connected to their local community.

"What I like most is helping people do something that they enjoy" Read how we helped this volunteer get work as a peer support worker!"

Thanking volunteers

A subset of Tweets in this topic are from organisations, thanking volunteers who have assisted them with their work. Many of these Tweets are linked to events such as Volunteers Week in June, which is 'a chance to celebrate and say thank you for the fantastic contribution millions of volunteers make across the UK'. The themes represented in these Tweets have some overlap with the 'messages of thanks' topic.

"Happy Volunteers' Week to all the volunteers who work in office admin roles. You keep everything running - thank you."

Personal experiences of volunteering

Finally, some Tweets were from individuals who spoke positively about their own experiences of volunteering across a range of activities, and the impact that it had on them.

"Months later I still miss being at [neighbourhood] adventure playground. I can't explain how lucky I was to work on the programme with these amazing children and young people!"

Mentions of voluntary and community organisations and activities were particularly frequent in Greenwich (where 8% of relevant Tweets referred to these topics), Barking and Dagenham (6%), and Newham (6%). Mentions were particularly low in Westminster (2%), Richmond upon Thames (2%) and Kensington and Chelsea (2%).

Turning to look at this topic by gender, males and females were similarly likely to engage with this topic on Twitter (3% of males and 4% of females mentioned the topic)[13]. Analysing the topic by age shows that mentions of voluntary and community activities accounted for 3% of Tweets from those aged under 18, 6% of those aged 18-29 and 5% of those aged 30 or over.

[13] Analysis by gender and age should be treated as indicative only due to the very low availability of gender and age information available in the dataset.

19-013606-01| Version 1 | Internal and client use only | This work was carried out in accordance with the requirements of the international quality standard for Market Research, ISO 20252, and with the Ipsos MORI Terms and Conditions which can be found at http://www.ipsos-mori.com/terms. © Greater London Authority 2019

1.8.3 Art and cultural activities

Around one in eight (12%) of the relevant Tweets addressed the topics of art and cultural activities. Over one third (37%) of these Tweets displayed a positive sentiment, while fewer than 1% had a negative sentiment. The word cloud below illustrates the words used in Tweets relating to art and cultural activities:

Tweets within this topic covered a very broad range of themes relating to creative events including art, literary, theatrical, cultural and musical events.

Exhibitions

A large subset of the Tweets mentioned exhibitions on a wide variety of subjects, many of which celebrated culture, heritage and history. Examples include the History of Afro-Caribbean Music, Muslim Heroes of the First World War, Street Art, and Women in Science and Medicine.

Tweets about exhibitions included promotional tweets posted by the organisations running the exhibition, or related organisations. Other Tweets were from individuals, sharing their experiences of visiting an exhibition.

"Excellent exhibition at [neighbourhood] library bringing back memories of [borough]. Fantastic work fighting inequality and racism"

"Great to celebrate #BlackHistoryMonth with colleagues. Also a fascinating Black History photography exhibition in the [borough] Town Hall."

Users commented on or advertised creative displays, including art created by children and local artists displayed in public spaces, and creative activities for people to get involved with.

Libraries

There were also many mentions of libraries within this topic. These were mainly in relation to events being held in libraries; including exhibitions, talks, workshops and reading groups. As with exhibitions, there was a combination of Tweets from organisations and from individuals who had experienced events at the libraries.

"Great fun was had at [neighbourhood] Library yesterday when young people entertained children visiting the library with stories and crafts!"

"I've visited [neighbourhood] Library to do some writing. What a wonderful library, with nice view, lovely kid's area with a good space for events. (Also very big slices of cake!)"

Festivals

Another group of Tweets within this topic centred around festivals of a range of forms. These included storytelling and craft festivals for children, art festivals, film festivals, music festivals (including small local music festivals) and comedy festivals among others. Some of these festivals were targeted at specific groups including young families, ethnic minorities, and those with disabilities.

"History & Heritage exhibitions, Arts & Crafts, Sports Games, World Cuisine, Kiddies Rides, Live Music, Poetry Reading, Beer Bar...there'll be something for everybody at [local festival] and we'll greet you with a smile"

"Are you ready for the summer? Check out the [local festival] Programme! [local festival] celebrates disabled and non-disabled artists in an exciting programme of music, dance and workshops!"

Diversity and inclusion

A theme that emerged throughout many Tweets relating to art and cultural activities was diversity. Tweets referred to many types of diversity, such as different nationalities and backgrounds, age, disabilities, gender or sexual orientation. Some Tweets mentioned events aimed at bringing together diverse groups of people in the local community. Other Tweets included positive comments about the value of multiculturalism.

"This festival looks like a great change to discover ways we can all help to create a dementia friendly society."

"We will be holding a multi-cultural Iftar (communal breaking of the fast) at the mosque next week. #Ramadan #MasjidRamadan"

19-013606-01| Version 1 | Internal and client use only | This work was carried out in accordance with the requirements of the international quality standard for Market Research, ISO 20252, and with the Ipsos MORI Terms and Conditions which can be found at http://www.ipsos-mori.com/terms. © Greater London Authority 2019

The theme of diversity emerged across Tweets about community activities and displays as mentioned above, with some activities aimed at raising awareness and educating the public about certain issues relating to inclusion and diversity.

"botanical gardens celebrate biological diversity - and the diversity of those who work within them! @PrideinLondon was a special day for [organisation]"

Mentions of art and cultural activities were particularly frequent in Greenwich (where 20% of relevant Tweets referred to these topics), Bexley (20%), Barking and Dagenham (17%), City of London (17%), Newham (17%) and Waltham Forest (17%). Mentions of this topic were relatively low in Hillingdon (7%).

Turning to look at this topic by gender, males and females were equally likely to discuss this topic on Twitter (8% of males and 9% of females Tweeted about this topic)[14]. Analysing the topic by age shows that the proportion of Twitter users mentioning this topic was similar across those aged under 18 (12% mentioned this topic), those aged 18-29 (15%) and those aged 30 or older (13%).

[14] Analysis by gender and age should be treated as indicative only due to the very low availability of gender and age information available in the dataset.

19-013606-01| Version 1 | Internal and client use only | This work was carried out in accordance with the requirements of the international quality standard for Market Research, ISO 20252, and with the Ipsos MORI Terms and Conditions which can be found at http://www.ipsos-mori.com/terms. © Greater London Authority 2019

1.8.4 Messages of thanks

Around one in ten (9%) of the relevant Tweets were offering messages of thanks. Around four in ten (38%) of these Tweets were identified as displaying a positive sentiment, while the remaining Tweets in this category (62%) were identified as having a neutral sentiment. The word cloud below illustrates the words used in the messages of thanks that the topic model identified:

There are a broad range of reasons for this gratitude; though many relate to donations, volunteering, lending support, or acts of kindness.

Thanks for support and donations

A subset of these Tweets thanked people for offering their support. These included support during sporting events such as marathons or football matches.

"Thank you to all my fellow runners and [borough] residents today. Very grateful for the support! A wonderful weekend of fitness and fun!"

Other Tweets thanked people for donations to a wide variety of charitable causes including those supporting disabled people, people with mental health issues, people in food crisis and homeless people, among others. As well as monetary donations, thanks were given for donations including food, clothing, toiletries and raffle prizes.

"We would like to thank [Twitter handle] for their generous donation, it has helped feed many less fortunate families and people here in [borough]"

Thanks for volunteering

Another subset of Tweets offered thanks to volunteers for giving their time to support charitable, community and faith groups. For example, volunteers assisting in charity shops, teaching older people to use technology, and providing first aid at community events.

"Thank you to [school] for the absolutely wonderful Air Cadets who attended our Veterans dinner today. They were all amazing, mixed so well with the veterans and were a brilliant to have. They did themselves and their school credit"

"In [neighbourhood], Jake has been getting Julie set up on her computer. A huge thank you to all our volunteers from [user handle] who spent their day helping local residents to email, text and watch videos."

Thanks to charitable organisations

Some Tweets gave thanks to a range of charitable organisations for the support that they had provided.

"Had three month follow up with [charity] this morning. Many thanks for this and your wider work in [borough] with support for stroke victims."

Thanks to public service staff

Finally, a number of Tweets gave thanks for acts of kindness from public service staff including the staff of Transport for London and the Metropolitan Police.

"@TfL thanks to the 2 awesome bus drivers on route 23 who helped me to get my bag left on one of them. Loving and caring attitude. Love you"

"Son and friends got attacked in [district]. Were able to get away on bus and called cops. Police were totally brilliant - came and took statements. Thank you to our fantastic, underfunded, police service"

Messages of thanks were particularly frequent in Harrow (where 17% of relevant Tweets referred to these topics), Bexley (15%), and Kingston-upon-Thames (14%). Messages of thanks were relatively infrequently sent from Tower Hamlets (6%), Kensington and Chelsea (6%) and Westminster (6%).

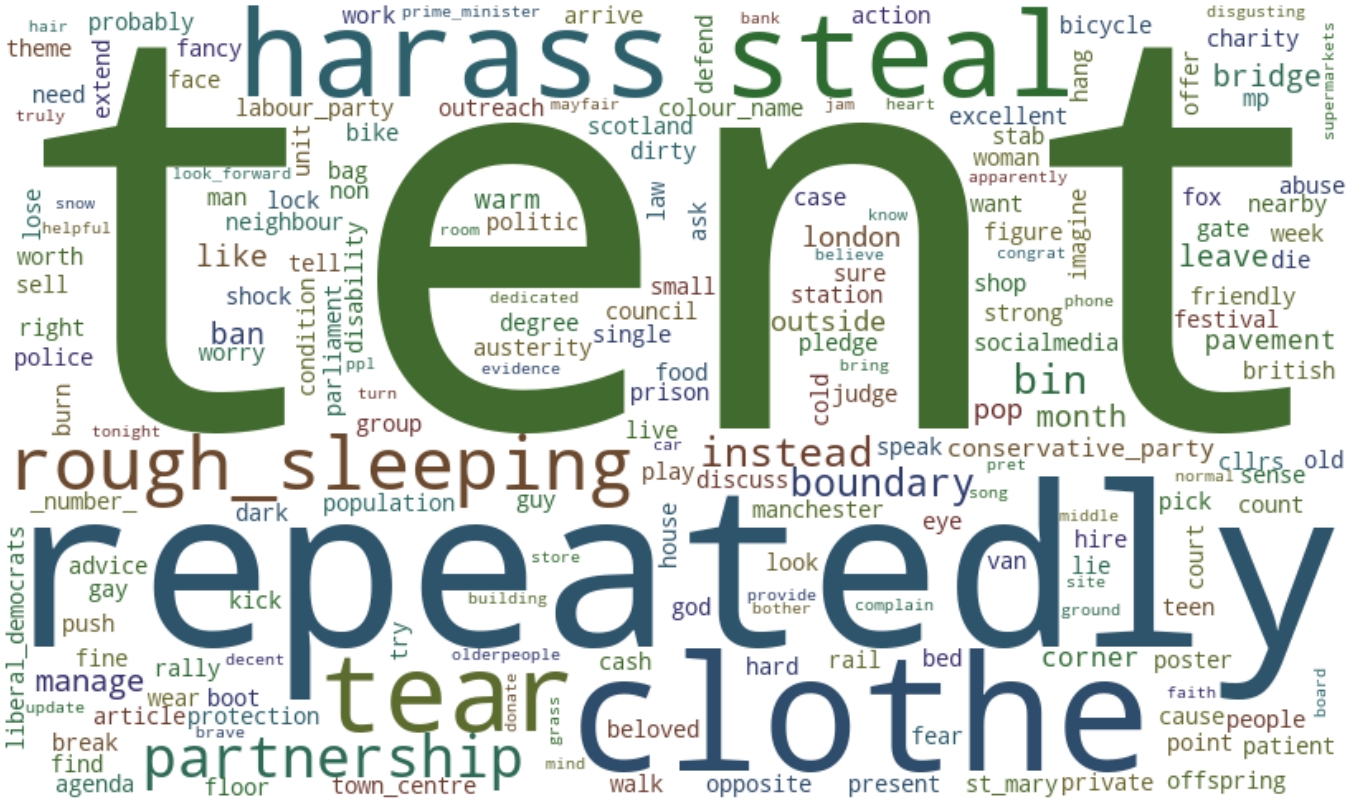
Turning to look at this topic by gender, a higher proportion of females (10%) than males (6%) sent messages of thanks on Twitter[15]. Analysing the topic by age shows messages of thanks account for 9% of relevant Tweets from those aged under 18, 8% of those aged 18-29 and 11% of those aged 30 or older. This indicates that this topic is discussed with similar frequency across the age bands.

[15] Analysis by gender and age should be treated as indicative only due to the very low availability of gender and age information available in the dataset.

1.9 Equality

1.9.1 Homelessness

Four percent of the Tweets were discussing issues relating to rough-sleeping and homelessness. The word cloud below illustrates the words used in Tweets relating to this topic.



This topic covered Twitter users' opinions towards homelessness in London. Posts often expressed empathy towards their situation, but also anger and concern about the growing number of homeless people in London. This is reflected in 25% of topic posts containing negative sentiment. (1% were positive and 73% were neutral.) For example, some of the Tweets were drawing attention to tents and other belongings being vandalised, stolen or taken down. Twitter users also recognised the need for homeless shelters, charity interventions and outreach programmes. While some Tweets are calling on their council to try to find ways to improve the problem, others are expressing disgust and shock. At the end of 2018 several articles were released suggesting local authorities were routinely taking down homeless people's tents. This explains the frequent usage of the word, 'repeatedly'.

Outreach initiatives and charity work

Twitter users recognised there are many ways in which they can help improve homeless people's daily wellbeing and long-term life outcomes. All over London charities are encouraging people to get involved in different initiatives, whether this be at the Ealing Soup Kitchen, providing free haircuts to help maintain health and hygiene, or engaging homeless people in games and group activities.

19-013606-01| Version 1 | Internal and client use only | This work was carried out in accordance with the requirements of the international quality standard for Market Research, ISO 20252, and with the Ipsos MORI Terms and Conditions which can be found at http://www.ipsos-mori.com/terms. © Greater London Authority 2019

"Huge thumbs up to Jack who volunteers every week offering the homeless in [borough] free haircut and a shave. There are people who care about the homeless."

Call for political action

A lot of the posts are appealing to authority figures and organisations. Many identify the lack of affordable housing, or a lack of government prioritisation, as the crux of the problem. Others speak to the multifaceted nature of the issue, and comment on how homelessness might be due to, or lead to, mental health difficulties. Viewed together the Tweets in this topic capture the complexity of the issue.

"1,732 children had meals from [borough] foodbanks last year, 2,600 are homeless and 79 families with children are in B&Bs, yet Tories talk about giving children the "best possible start to life".

Concern for the homeless

A subset of Tweets expressed concerns about how homeless people are treated by others and about how their situation might improve. In some instances, Tweets referenced recent news stories, for example, the story of a homeless man spending five nights on the streets after being discharged from hospital, to evidence their opinion and raise awareness of the severity of the problem.

"Whilst [coffee chain] in USA has opened its doors to 'everyone', a homeless person taking a nap in the [neighbourhood] branch has just been kicked out."

Interactions with the homeless.

Some Tweets were commenting on positive and negative social exchanges with homeless people. Some users were sharing first hand interactions and observations, others were giving their opinion on second-hand interactions.

"Shocked to watch [supermarket] harass a homeless guy sitting outside their premises and tell him he was "disgusting""

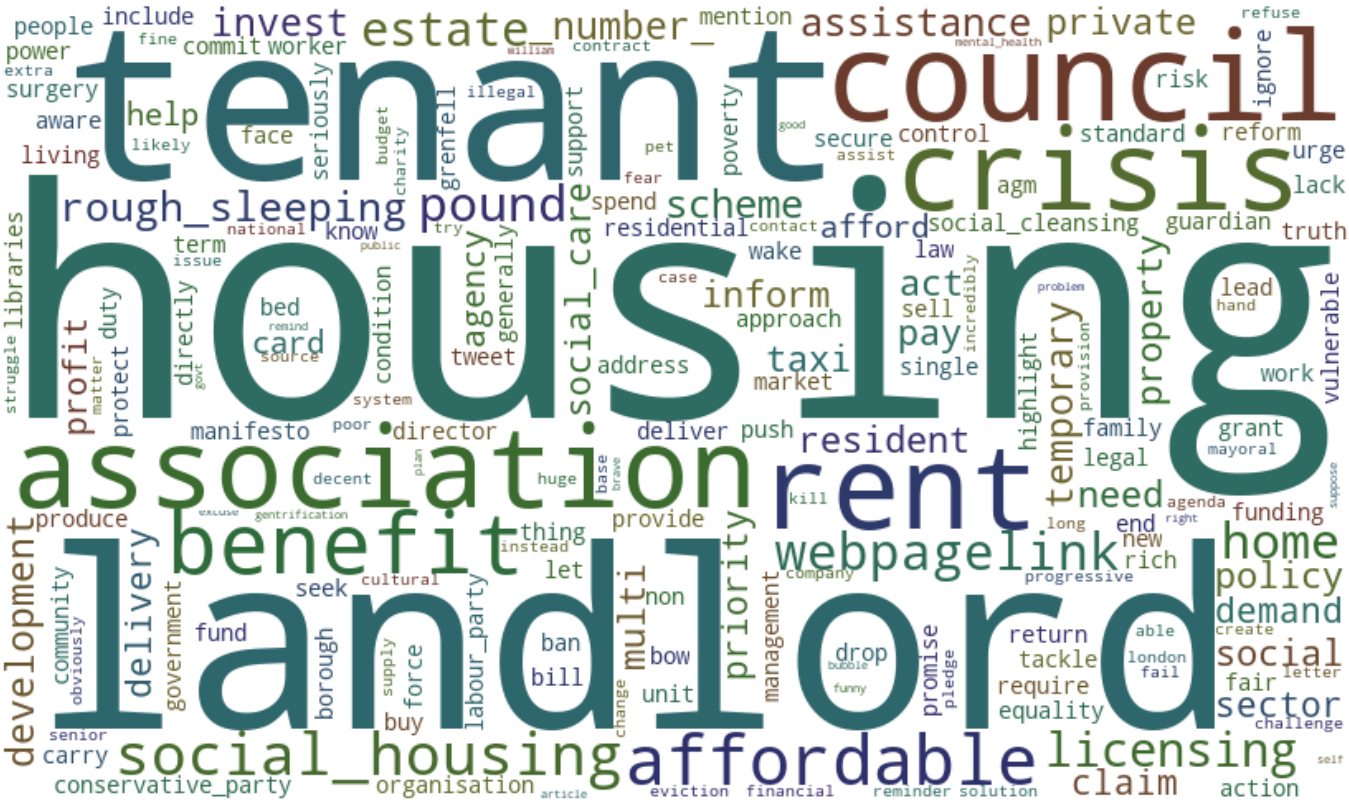
Tweets relating to homelessness were particularly frequent in Newham (where 11% of relevant Tweets referred to this topic), Islington (8%), Lewisham (8%) and Camden (6%). Among users where gender was known, males and females were equally likely to Tweet about homelessness (mentioned 3% of both males and females)[16]. Analysing the topic by age shows that messages of thanks account for 3% of relevant Tweets from those aged under 18, 2% of those aged 18-29 and 4% of those aged 30 or older. This indicates that this topic is discussed with similar frequency across the age bands.

[16] Analysis by gender and age should be treated as indicative only due to the very low availability of gender and age information available in the dataset.

19-013606-01| Version 1 | Internal and client use only | This work was carried out in accordance with the requirements of the international quality standard for Market Research, ISO 20252, and with the Ipsos MORI Terms and Conditions which can be found at http://www.ipsos-mori.com/terms. © Greater London Authority 2019

1.9.2 Housing

Six per cent of the relevant Tweets concern issues relating to housing. Just one in five (21%) of these Tweets displayed a positive sentiment, while two in three (67%) had a negative sentiment. While the sentiment analysis should be treated with caution, it is worth noting that the proportion of negative sentiment was particularly high at this topic. The word cloud below illustrates the words used in Tweets relating to housing:



The Tweets contained in this topic cover a range of themes related to social housing, affordable homes, private renting, homelessness, and redevelopment.

Social housing

Regarding social housing, many of the Tweets highlighted the need for more social housing across London, and within specific areas of London. Some Tweets also discussed the removal or regeneration of social housing with some linking this to regeneration. There are calls to better ensure that residents are properly consulted before decisions about regeneration are made.

"[borough council] didn't bother consulting with residents of [housing estate] before putting their homes forward for regeneration."

"It's really sad. The multicultural, diverse [borough] is slowly becoming a soulless and generic place for middle class people who can afford to pay the ridiculous prices of the 'apartments' that are being built instead of social housing."

Some Tweets mentioned the number of homeless people and the need for additional support to help homeless people find secure housing.

"[borough] is suffering from the housing crisis. There are 10,000 waiting for housing and 2,000 homeless households. Check out [Twitter handle]'s plan to tackle the housing crisis and build the homes we need"

Affordable homes

Tweets also raised the need for more affordable homes, and the need to ensure that all new developments include affordable homes. There were also concerns that many new developments do not meet planning requirements in terms of the number of affordable homes that are built.

"Despite London's housing crisis, half billion-pound development beside [park] won't include a single affordable flat."

"Planning consent for this development required 76 social rented homes; only 45 were provided"

Some Tweets praised what they saw as good examples of where councils have prioritised affordable and social housing. Others criticised what they viewed as over-development of certain areas of London.

"When people ask what sort of housing we need, just one of the places I mention is [borough] & the [housing development]. Completely affordable housing in [neighbourhood], dozens for older people with a care needs"

"Council approves 250 new homes. This is madness! [neighbourhood] and surroundings are already hugely over developed! [borough council] have spoiled what was once a lovely borough to live in!!"

Learnings from Grenfell

A subset of Tweets referred to the learning from Grenfell and actions that need to be taken in the future. There are also a series of Tweets about the importance of ensuring that Councils prioritise safeguarding against another such tragedy, for example by replacing fire doors in social housing where necessary.

"Honest discussion on public housing post #Grenfell and standing ovations for [union]. Thank you to all members on the ground at Grenfell"

Private rents

Several Tweets highlighted concerns about the cost of private rents in London and the practices of private landlords, including illegal evictions

19-013606-01| Version 1 | Internal and client use only | This work was carried out in accordance with the requirements of the international quality standard for Market Research, ISO 20252, and with the Ipsos MORI Terms and Conditions which can be found at http://www.ipsos-mori.com/terms. © Greater London Authority 2019

"Sadly, it is not unusual to see landlords subletting illegally to immigrants relying mostly on their lack of knowledge about their rights."

Tweets relating to housing were particularly frequent in Newham (where 10% of relevant Tweets referred to housing), Hammersmith and Fulham (9%), Haringey (9%) and Kensington and Chelsea. Tweets relating to housing were less likely to refer to the City of London (1%), Bromley (2%) and Richmond upon Thames (3%).

Turning to look at this topic by gender, males were more likely than females to engage with this topic on Twitter (6% of males mentioned this topic compared to 3% of females)[17]. Analysing the topic by age shows Tweets relating to housing account for 5% of relevant Tweets from those aged under 18, just 3% of those aged 18-29 and 6% of those aged 30 or older.

1.9.3 LGBT rights

Three percent of Tweets were associated with LGBT rights and related issues. The word cloud below illustrates the words used in Tweets relating to the topic:

This topic primarily comprised of Tweets that celebrated the LGBT community and LGBT social movements. Many Tweets supported LGBT events, inclusivity and diversity - this was reflected in 38% of topic posts containing positive sentiment (61% were neutral).

[17] Analysis by gender and age should be treated as indicative only due to the very low availability of gender and age information available in the dataset.

By taking a bottom-up approach to data analysis we expect topics to contain differing degrees of noise. In other words, some Tweets will sit more appropriately under a topic than others. As seen here, the word 'flag' appears most frequently in this topic. This occurred because the query did not always distinguish the use of the word within the context of Pride, to its use in other contexts (i.e. other types of flags, e.g. the England flag). With this in mind, Tweets often related to one of the following themes:

Displays of the Pride flag

A few Twitter users mentioned how they were, or had seen others, displaying the pride flag in solidarity with the LGBT movement, sometimes in relation to the parade.

"Today's the day! Our #rainbow flag is being raised at [Twitter handle] HQ in [neighbourhood] in readiness for the @PrideInLondon parade on Saturday #PrideMatters #PrideInLondon"

"[Twitter handle] - Great to see the Pride Flag on display at your site"

Within this theme we see some overlap with the 'messages of thanks' topic. As seen in the example below, many posts are expressing their appreciation and happiness about others proudly raising Pride flags.

"This morning the rainbow flag will fly over the town hall for #Pride2018. A first for [borough] but hopefully every year from now on. #PrideMatters"

Support and celebration of LGBT+ rights

In a selection of Tweets, Twitter users demonstrated support for the LGBT community and celebrated progress being made. The Tweets included in this topic speak of the projects being run by organisations within London.

"Meet some of the team behind the [borough] #LGBTQ+ Forum. A great bunch of dedicated volunteers who serve their community all year"

Tweets also demonstrated support people are giving on an individual level, for example, as LGBT+ allies. Tweets openly support equal rights and gender equality, and challenge phobias and discrimination.

"Do not insult lgbtq like you have not come across gay/trans/queer people everyday of your life."

"Proud to have been an ally of the LGBT community for decades."

19-013606-01| Version 1 | Internal and client use only | This work was carried out in accordance with the requirements of the international quality standard for Market Research, ISO 20252, and with the Ipsos MORI Terms and Conditions which can be found at http://www.ipsos-mori.com/terms. © Greater London Authority 2019

LGBT events

A wide range of LGBT related events were covered in these Tweets. They included LGBT dance nights and monthly poetry events. Tweets express a desire to see the opinions of those within the LGBT community reflected in all aspects of society, whether this be political, educational or creative.

"We're just a small group of volunteers who give up our free time to try and improve things for LGBT+ people locally. We are not campaigners but we will stand up for things we think are wrong or should be challenged."

A large selection of these Tweets referred to the annual London Pride Parade, and discussed the range of music, dance and food on offer. Organisations also promote how they will be participating on the day.

"[university] LGBTQ+ society joined this year's #LondonPride parade! Here is a preview of our official Pride video."

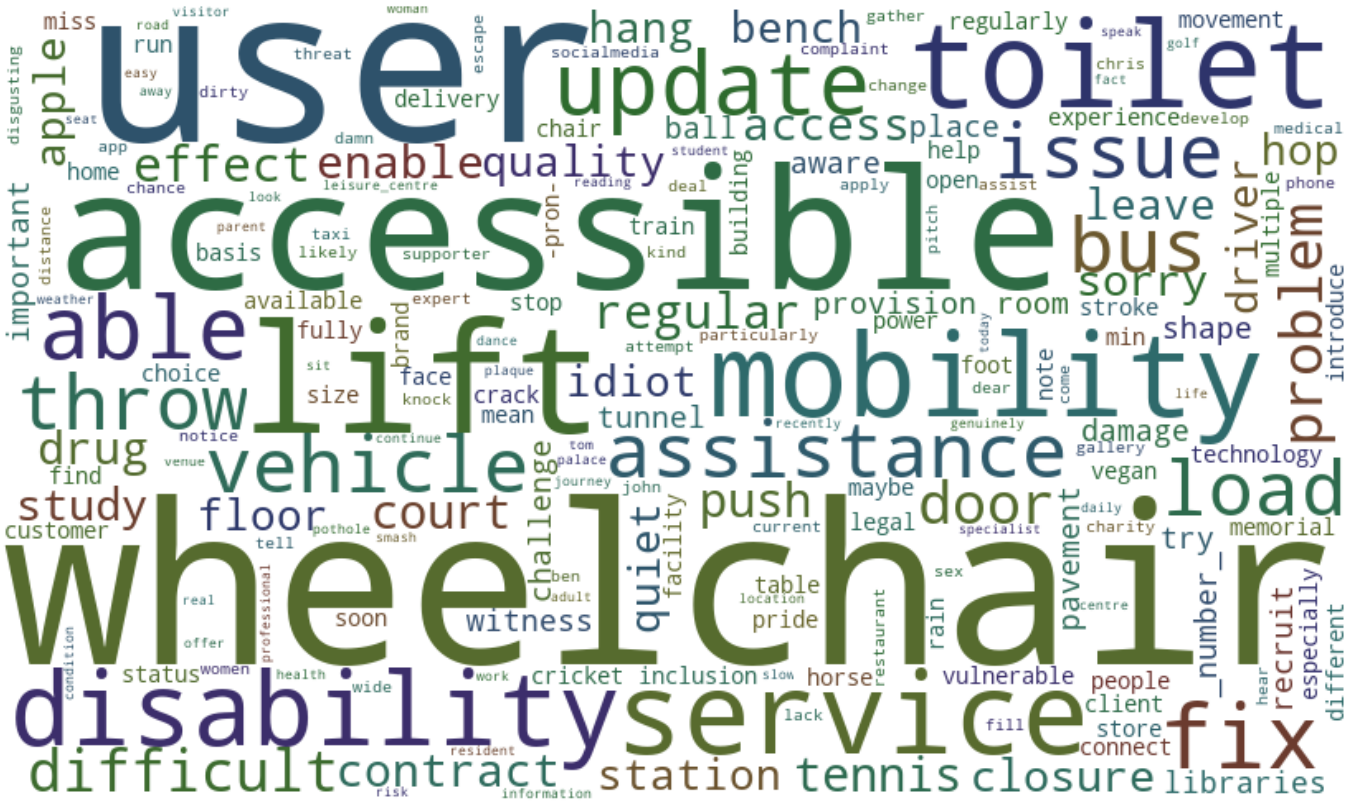
Tweets relating to LGBT rights were particularly frequent in Havering (where 4% of relevant Tweets referred to this topic) and Lambeth (4%). Turning to look at this topic by gender, males and females were equally likely to post Tweets relating to LGBT rights (3% of both males and females Tweeted about this topic)[18]. Analysing the topic by age shows Tweets relating to LGBT rights for 3% of relevant Tweets from those aged under 18, 4% of those aged 18-29 and 3% of those aged 30 or older. This indicates that this topic is discussed with similar frequency across the age bands.

[18] Analysis by gender and age should be treated as indicative only due to the very low availability of gender and age information available in the dataset.

19-013606-01| Version 1 | Internal and client use only | This work was carried out in accordance with the requirements of the international quality standard for Market Research, ISO 20252, and with the Ipsos MORI Terms and Conditions which can be found at http://www.ipsos-mori.com/terms. © Greater London Authority 2019

1.9.4 Accessibility

One percent of the Tweets were discussing disabled accessibility to different services and environments. Compared with other topics, a relatively high proportion of Tweets within this topic were categorised as having negative sentiment (2%), although a higher proportion had positive sentiment (29%). The word cloud illustrates the words used in Tweets relating to this topic.



As displayed in the word cloud, most Tweets commented on wheelchair accessibility.

Accessible public transport

This was frequently discussed in the context of public transport; particularly trains and buses. Posts discussed difficulties faced by disabled people at all stages of the travel journey. Some highlighted negative experiences of public transport whereas others were about difficulties that had arisen before the travel journey had begun, such as trying to access the station toilets or lifts. For example, one Tweet referred to the lack of wheelchair accessible tables at a coffee shop at a major Underground and Rail station. The lack of public toilets in London, whether this be wheelchair accessibility, at stations or otherwise, appeared as a common topic theme. Other Tweets were signposting accessibility problems more generally. For example, accessing public places e.g. car parks, traffic and road works.

The tone of the Tweets within this topic was generally that of an announcement or complaint. Tweets were often directed at local government bodies.

Tweets could be split into one of the two themes below:

Accessible public toilets

Some Tweets highlighted the lack of public toilets to the facility provider. This was most commonly in the context of there being too few designed to accommodate people with reduced mobility.

"In the [borough] they insist on accessibility with licensing. It's not unreasonable to expect bigger venues to provide toilets"

"[university] doesn't have running water in accessible toilet, reception send people to the regular toilets. No gender-neutral access"

"There are people sat in the toilet on the 14.14 from [rail station]. It's impossible to access in a wheelchair or if you had incontinence."

Treatment of those with disabilities

From the Tweets it would suggest that in London some people experience unequal access to transport, libraries, restaurants and bars because of their physical impairments. In the first Tweet below, someone is referring to a story in the Evening Standard about a blind man being told that he would not be able to eat inside the restaurant because of his guide dog.

"Shame. Your first restaurant in [neighbourhood] was the only place in [neighbourhood] with a disabled toilet. Your staff need training and this gent needs free food for life."

"Just saw yet another bus driver leave a wheelchair user behind because of problems with the ramp - come on [Twitter handle] please try harder"

Tweets relating to accessibility were particularly frequent in Havering (where 4% of relevant Tweets referred to this topic) and Lambeth (4%). Turning to look at this topic by gender, males and females were equally likely to post Tweets relating to accessibility (3% of both males and females mentioned this topic)[19]. Analysing the topic by age shows Tweets relating to LGBT rights for 3% of relevant Tweets from those aged under 18, 4% of those aged 18-29 and 3% of those aged 30 or older. This indicates that this topic is discussed with similar frequency across the age bands.

[19] Analysis by gender and age should be treated as indicative only due to the very low availability of gender and age information available in the dataset.

19-013606-01| Version 1 | Internal and client use only | This work was carried out in accordance with the requirements of the international quality standard for Market Research, ISO 20252, and with the Ipsos MORI Terms and Conditions which can be found at http://www.ipsos-mori.com/terms. © Greater London Authority 2019

Top-down analysis

The bottom-up topic modelling succeeded in identifying many of the key elements of social integration within the cleaned data corpus. Notably, there were substantial quantities of topics relating to participation. However, it was notable that some topics of interest were missing from the topic model; particularly those relating to relationships.

There are several possible reasons why these topics may have been missing from the topic model. The first is that these topics are not being discussed on Twitter. The second possible reason is that the topics are being discussed on Twitter, but the search query did not collect Tweets pertaining to these topics (and that they are therefore not represented in the data corpus). This possibility is considered in the 'learning and considerations' chapter. The third possibility is that the topics are present in the data corpus, but the topic model did not successfully identify them. The topic model may have failed to identify the topics because either they are not present in high enough quantities within the cleaned data corpus, or because the language used is too diverse for the topic model to recognise them as belonging to a cohesive group.

To validate the topic model, and check whether this data is present in the data corpus, we conducted top-down analysis on key topics relating to social integration which were not present in the topic model:

- Helping neighbours
- School activities and events
- Conversations with friends
- Local area change and affordability (this might include shops closing and new housing developments)

To conduct this analysis, a Boolean search query was written, containing key terms relating to the topic of interest. This query was then run on the cleaned data corpus of 50,039 Tweets and refined further as necessary. The findings from this analysis are discussed in this chapter.

1.10Helping neighbours

A search query was developed to identify users within London talking about helping neighbours. Some topics associated with helping neighbours are already present in existing topics. For example, Tweets from small organisations or local groups that aim to get residents together, reduce social isolation or tackle particular issues came out in the topics of loneliness and social isolation and art and cultural activities. Other relevant topics however, such as helping neighbours at an individual level, were not explicitly captured within the existing topics.

19-013606-01| Version 1 | Internal and client use only | This work was carried out in accordance with the requirements of the international quality standard for Market Research, ISO 20252, and with the Ipsos MORI Terms and Conditions which can be found at http://www.ipsos-mori.com/terms. © Greater London Authority 2019

As well as identifying explicit mentions of helping neighbours, the search query sought to identify mentions of favours, locals, community centres, loneliness and old age. The full search query used can be found in the appendices.

Neighbours improving their area/community

A number of users Tweeted about being involved in group activities with their neighbours and other locals. Examples included more formal groups working together to make their neighbourhood safer, as well as residents getting together to do some gardening or to rescue ducks.

"It was great to join the community gardening event with [user handle] and my children today. Residents getting together to care about this lovely green space at the heart of [borough] and helping each other"

"This evening's duck rescue in [neighbourhood] not my usual flock! Thanks to my neighbours for all your help returning them to the river!"

"Community members sharing and working towards solutions for youth violence in [borough] at our Safer Neighbourhood Board Meeting"

Promoting a community culture of helping neighbours

Some Tweets related to this topic came from local groups or small organisations that aimed to promote communities in which people helped each other. These groups sometimes talked about events they were organising to get locals together, or referred to specific places where locals could go to meet each other.

"It's Community Centre week! We are celebrating neighbourhoods & neighbourliness. Our neighbourhood is [neighbourhood] but we are open to our neighbours all over"

"[borough] helps neighbours to create a community where they look out and care for each other."

Individual examples of helping neighbours

The search query for this topic also tried to uncover examples of Tweets that mention helping neighbours at an individual level. The Tweets about this were limited and referred to very specific individual situations, such as helping to raise awareness of a neighbour's cat that had gone missing. Some examples are shown below.

"My next-door neighbour (a few doors up from me) in South East London played in an orchestra. I learned about him when I went to help his wife when he was terminally ill a year ago."

"If anyone comes across Patch in [local area] please call, she's our neighbour's cat and their daughter is very upset."

19-013606-01| Version 1 | Internal and client use only | This work was carried out in accordance with the requirements of the international quality standard for Market Research, ISO 20252, and with the Ipsos MORI Terms and Conditions which can be found at http://www.ipsos-mori.com/terms. © Greater London Authority 2019

"In [neighbourhood], volunteers from [organisation] are supporting their older neighbours to get online, all while having a great natter."

1.11School activities and events

A specific search query was created for school activities and events. We found that school activities and events were often discussed in the context of charity events, community funding and local projects, and so it was interesting to explore what themes emerged when these subject matters were analysed together, under this predefined topic query.

Broadly this topic covered the many ways in which individuals, organisations and educational institutions were trying to improve children's overall educational experience. Twitter users are not just concerned with the quality of teaching, they often want to know whether children are breathing clean air and eating healthy food. Many of the Tweets contained messages of thanks to various organisations and clubs for volunteering resources and time to London schools. From the data three main themes were identified.

Educational events

A number of Tweets highlighted educational talks and events that were held within schools for pupils. Topics covered included human rights, LGBT identities, road safety, financial management, and religious festivals. In many cases the Tweets sharing information about these events were posted by the school or by the organisation running the event.

"Thanks to [school] for hosting human rights talks. Pupils asked terrific questions."

"Two members of the bank have been coming into school to talk to Year 5 about the topics such as banking, saving money and budgeting. The aim of the workshops has been to teach children about money and saving in a fun way."

Fundraising and local giving

Many groups and individuals are finding ways to improve the schools in their local area. A selection of the Tweets discussed how individuals were raising money for school equipment.

"My husband loves golf - but a golf marathon may not be much fun! Worthwhile to raise money for [school] to buy facilities for their pupils with complex needs though"

Many organisations have found different and sometimes unique ways of volunteering and giving back to the community. In some instances, this was in the form of monetary giving.

"Thank you [organisation] for the lovely donation to the school fair!"

But in many instances people were volunteering their time and skills, sometimes in ways that might help improve children's opportunities and life skills after leaving school.

19-013606-01| Version 1 | Internal and client use only | This work was carried out in accordance with the requirements of the international quality standard for Market Research, ISO 20252, and with the Ipsos MORI Terms and Conditions which can be found at http://www.ipsos-mori.com/terms. © Greater London Authority 2019

"Volunteers from [organisation] are helping [school] prepare for mural paintings, to brighten up the playground for children"

Air quality around schools

This was not only a strong but interesting finding as local air quality was a theme that emerged within the participation topic. An increased public awareness about the detrimental effects of air pollution is echoed in concerned Tweets.

"Took my children on holiday and they remarked how they could no longer 'smell pollution in the air' - their school is right by a main road in [borough] and our kids are being slowly poisoned"

Support of the steps being taken to help schools fight toxic air in London through the Mayor's Air Quality Audit Programme[20] and Greener City Fund was found in a collection of the Tweets and in some respects, overlaps with the previous 'fundraising and local giving' theme.

"@MayorofLondon well done on your great work with school's air quality, just so you know we have already installed Green Screens at [school]."

"Four schools in [borough] are receiving funds from @SadiqKhan to protect pupils from London's pollution. The Mayor is playing his part, now it is the Govt's turn to help reduce London's air pollution."

1.12Conversations with friends

The GLA were interested in understanding conversations that take place on Twitter between friends. Identifying these conversations is challenging as Twitter provides no way to distinguish between accounts belonging to private individuals, public figures, and organisations. As such, the top-down analysis aimed to identify conversations with and about friends by searching for occurrences of words associated with friendship. As discussed in the 'learning and considerations' chapter, further research on approaches to make these distinctions between types of accounts could be beneficial, for example for identifying conversations between individuals about topics relevant to social integration that do not explicitly mention words associated with friendship.

Very few, if any, conversations between friends were identified and a limited number of Tweets mentioning friends were identified. A small subset of Tweets was sent by Twitter users on behalf of their friends. For example, raising awareness of crimes that had been committed, asking for sponsorship for friends' fundraising activities, or encouraging participation in team activities.

"Our brother, friend and bandmate had his pedals and audio recorder stolen last night in the [neighbourhood] last night."

[20] https://www.london.gov.uk/what-we-do/environment/pollution-and-air-quality/mayors-school-air-quality-audit-programme

"My friend needs some team members for a game of football tomorrow evening. 8 a side. 2 hours. [neighbourhood]."

"A friend's family raise money for a local cancer charity each year. This time someone broke in & stole money from a safe. If you know anything (happened in [borough]) please tell police"

"A friend of mine that lives in [borough] organises a run every Sunday morning do you want to participate?"

Other Tweets were about experiences that Twitter users had shared with their friends. These included sporting, cultural, and religious activities, such as Eid celebrations.

"Beautiful day to celebrate Eid in our stunning [park] with over 1000 Muslims from all around [neighbourhood] gathered for prayers and worship."

1.13Local area change and affordability

A search query was developed to identify discussion related to local area affordability and neighbourhood change within London. Many topics associated with neighbourhood change and affordability are already evident across the existing topics. For example, Tweets within the housing topic include discussion of the need for more social housing and affordable housing, and Tweets within the ethnic minorities topic include discussion of neighbourhood change. However, other elements of these topics – including the closure of shops and more general neighbourhood changes - were not explicitly captured within the existing topics.

Therefore, the search query sought to identify mentions of closures, new developments, re-developments, unaffordability, being priced out, evictions, regeneration and social change. The full search query used can be found in the appendices.

Housing

Some of the identified Tweets voiced concerns about neighbourhood change in relation to increases in property and rental prices, which led to Londoners being priced out of areas where they had previously been resident.

"If you see [estate agent] you should worry. I recall seeing them appear in [neighbourhood] and then [neighbourhood] 10 years ago now you cannot afford to rent a bedsit there. There are upsides but it's pricing people out."

Linked to this, and as noted within the housing topic, there was concern about the demolition of the housing estates which are home to "ordinary Londoners". Some called for additional consultation with residents prior to decisions being made, and increased oversight to ensure that – where demolitions do take place – any re-development includes enough provision of affordable and/or social housing.

19-013606-01| Version 1 | Internal and client use only | This work was carried out in accordance with the requirements of the international quality standard for Market Research, ISO 20252, and with the Ipsos MORI Terms and Conditions which can be found at http://www.ipsos-mori.com/terms. © Greater London Authority 2019

"STOP SOCIAL CLEANSING AND HALT DEMOLITIONS [borough council]: Petition to demand a vote on the demolition of [borough]'s estates. - Sign the Petition!"

Some Tweets explicitly linked affordability of housing and neighbourhood change to 'social cleansing', suggesting that councils were favouring housing developments which were unaffordable for residents – and deprioritising affordable housing - in order to proactively change the demographic profile of the local population, both in terms of ethnicity and household income.

"They couldn't care less about those in need. Hello to social cleansing and goodbye to looking after the people who were born and bred in [borough]. They are in it purely for greed."

"One of the speakers is a friend and very concerned about social injustices, and about the social cleansing of [borough] by a [political party]-led council hellbent on changing the borough's demographic."

"Median household income per year for [borough] - about £32,000. Therefore, that flat is unaffordable to most residents. If that's not social cleansing, then what is it?"

Small businesses

A second area of concern relating to neighbourhood change centred on small local businesses which had been evicted due to re-developments, or otherwise damaged by the increased rents associated with regeneration. For example, concerns were raised that the long-standing traders in Elephant and Castle shopping centre would not be given enough protection under the current redevelopment plans. There were also several mentions of redevelopment in Peckham and the impact that this will have, and already has had, on local businesses. Some Tweets mention the increase in empty units within Peckham, which has resulted from increased rents.

"[shopping centre] demolition approved. There are concerns the proposals do not include enough affordable housing and that existing traders would not have enough protection."

"Inevitably some businesses affected by the work will have to relocate. This is why people have an issue with gentrification"

In some cases, Twitter users praised actions taken by individuals and councils to ensure that redevelopments meet the needs of current residents and that business are fully supported.

"[borough council] has reaffirmed position when it comes to new developments and the responsibility developers have to provide affordable housing and public facilities for community use. Very welcome indeed!"

19-013606-01| Version 1 | Internal and client use only | This work was carried out in accordance with the requirements of the international quality standard for Market Research, ISO 20252, and with the Ipsos MORI Terms and Conditions which can be found at http://www.ipsos-mori.com/terms. © Greater London Authority 2019

Considerations around representativeness

One of the important principles of social research is representativeness; that the attitudes that are studied recognisably reflect the views of a known group. Representativeness is not always important, but it is vital when the research aims to draw inferences about a wider population. For example, if the findings from this research were to be generalised to the broader population of London, representativeness is a crucial consideration.

Traditional, offline research ensures representativeness either by selecting research participants on the basis of them being a representative cross section of society based on a range of factors, or by controlling for these factors after data collection. These factors include age, gender and socio-economic status.

Using social demography, this chapter will explore the extent to which the Twitter dataset used for this research – and Twitter data more generally - is representative. It will seek to explore the following questions:

- How representative are Londoners who access Twitter of the wider London population?
- How representative was the dataset of the conversation taking place on Twitter?

To assess the proportion and demographic profile of Londoners who use Twitter, we used conventional social research methods to take offline measurements from a representative sample of Londoners. Specifically, data was collected across a series of face-to-face omnibus surveys between 2017 and 2019[21]. In these surveys, a representative sample of 545 Londoners aged 18 or over were asked 'Which of the following websites have you used or visited within the last three months?' and asked to select from a list of commonly used social media websites including Twitter.

1.14 How representative are Londoners who use Twitter of the wider London population?

In total, one in five Londoners (21%) report having visited or used Twitter within the last three months. Although this is a fairly small proportion of the population, this does not in itself confirm that the Londoners who use Twitter are not representative of Londoners more broadly. To understand this further, it is necessary to understand the demographic profile of the 21% who have accessed Twitter, and the remaining 79% who have not. It should be noted however that, although demographics are one indicator of the representativeness of a sample, there may be other differences between those who access Twitter and those who do not that are not so easily measured or controlled for.

[21] Data was collected in four waves; in February 2019, April 2018, February 2018 and November 2017. Data was weighted to age within gender, age within working status, tenure and ethnicity based on the 2017 ONS mid-year population estimates.

19-013606-01| Version 1 | Internal and client use only | This work was carried out in accordance with the requirements of the international quality standard for Market Research, ISO 20252, and with the Ipsos MORI Terms and Conditions which can be found at http://www.ipsos-mori.com/terms. © Greater London Authority 2019

1.14.1 Representativeness by age

Table 6.1 shows the proportion of Londoners in each age group who reported accessing Twitter in the last three months. Londoners aged 18-34 are significantly more likely to have accessed Twitter than those aged 35-54, and both these groups are more likely to have accessed Twitter than those aged 55+. This indicates that the content shared on Twitter is less likely to represent the views and experiences of older Londoners than it is younger Londoners.

Table 1.1: Proportion of Londoners who have used or visited Twitter in the last three months by age.

Age	18-34	35-54	55+	Total
Base	*(205)*	*(164)*	*(138)*	*(545)*
Accessed Twitter	30%	19%	11%	21%
Not accessed Twitter	70%	81%	89%	79%
Total	100%	100%	100%	100%

1.14.2 Representativeness by gender

Furthermore, although the proportions of male and female Londoners who report accessing Twitter are similar (22% and 19% respectively), there are some significant differences when the gender of Twitter users is analysed by age.

Table 6.2 shows that, although among Londoners aged 18-34 and 35-54 males and females are equally likely to have accessed Twitter, stark differences by gender emerge among those aged 55+. Within this age group, around one in six males (17%) have accessed Twitter, compared with fewer than one in twenty females (4%). Although this difference in gender for those aged 55+ is statistically significant, it is based on a small sample size, and is not reflected in data on Twitter usage across the UK[22]. As such it should be treated as indicative only. However, it highlights a risk that the views and experiences women aged 55+ - who constitute 15% of the population of London[23] – are largely absent from Twitter.

[22] Ipsos MORI Tech Tracker: https://www.ipsos.com/sites/default/files/ct/publication/documents/2017-08/Ipsos_Connect_TechTracker_Q2_2017_0.pdf
[23] Based on 2017 mid-year census estimates: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates

Table 1.2: Proportion of Londoners who have used or visited Twitter in the last three months by age and gender.

Age, Gender	18-34, Male	18-34, Female	35-54, Male	35-54, Female	55+, Male	55+, Female	Total
Base	*(103)*	*(102)*	*(102)*	*(101)*	*(64)*	*(72)*	*(545)*
Accessed Twitter	29%	31%	19%	19%	17%	4%	21%
Not accessed Twitter	71%	69%	81%	81%	83%	96%	79%
Total	100%	100%	100%	100%	100%	100%	100%

1.14.3 Representativeness by socio-economic status

Finally, the proportions of Londoners who report accessing Twitter vary by socio-economic status (as approximated by occupation). Table 6.3 demonstrates that Londoners in social grade AB[24] (26%) and those in social grade C1C2[25] (23%) are significantly more likely to have accessed Twitter than Londoners in social grade DE[26] (13%). Based on 2011 census data, one in four Londoners (23%) fall into social grade DE, suggesting that this is a large segment of the population whose views and experiences are underrepresented on Twitter.

Table 1.3: Proportion of Londoners who have used or visited Twitter in the last three months by social grade.

Social grade	AB	C1C2	DE	Total
Base	*(178)*	*(188)*	*(193)*	*(545)*
Accessed Twitter	26%	23%	13%	21%
Not accessed Twitter	74%	77%	87%	79%
Total	100%	100%	100%	100%

In summary, a relatively small proportion of Londoners (21%) have either used or visited Twitter within the last three months. Of these Twitter users, older populations – particularly females – and those in

[24] Social grade AB refers to those in higher and intermediate managerial, administrative, professional occupations. Based on 2011 census data, 28% of Londoners fall into the social grade AB.
[25] Social grade C1 refers to those in supervisory, clerical and junior managerial, administrative professional occupations while C2 refers to those in skilled manual occupations. Based on 211 census data, 49% of Londoners fall into either the C1 or C2 social grade.
[26] Social grade DE refers to those in semi-skilled & unskilled manual occupations, the unemployed and the lowest grade occupations. Based on 2011 census data, 23% of Londoners fall into social grade DE.

19-013606-01| Version 1 | Internal and client use only | This work was carried out in accordance with the requirements of the international quality standard for Market Research, ISO 20252, and with the Ipsos MORI Terms and Conditions which can be found at http://www.ipsos-mori.com/terms. © Greater London Authority 2019

lower social grades are underrepresented. It should be noted that these findings are based on the proportion of Londoners who either use *or* visit Twitter. The population that use Twitter is likely to be even more skewed than this data suggests.

While this doesn't discount the use of Twitter for meaningful social research, it limits the reliability with which any findings from Twitter can be generalised to the broader population of London. However, to a greater or lesser extent, the same considerations apply to whichever research methodology is used. Indeed, for some sectors of the population (e.g. young men), Twitter may be more representative than traditional research methods, which typically struggle to engage with some groups. Therefore, when used mindfully, and triangulated with other data sources, Twitter can still provide a valuable source of insight despite its limitations.

1.15 How representative was the dataset of the conversation taking place on Twitter?

A second question is the extent to which the datasets gathered for use in this research represent what is really being discussed on Twitter. Acquiring data on Twitter is different from sampling used in conventional attitudinal social research. As described in the methodology chapter of this report, the Twitter data to be exported is identified using a combination of a search query and filters. The data that is exported is therefore solely dependent on the precise key words which are included in the search query and the decisions that are made about which filters to use.

This form of acquiring Twitter data presents a problem, because decisions made by the researcher may mean that the dataset does not include Tweets that are relevant to the things being studied. The issue of missing relevant data is a difficult one. Because missing, it is difficult to know how much of it has not been collected, or what kind of data is indeed missing. However, given the methodology of the project, it is possible to identify the main causes of missing relevant data.

The choice of keywords included in the query

The primary cause of missing relevant data is the choice of keywords that are used when building the search query. The solution attempted for this project, was to use deliberately over-expansive keywords. Rather than using keywords associated with social integration issues, we used the full list of London boroughs, constituencies, Underground stations, Overground stations, National Rail stations and tram stations as well as references to major parks and town centre names (see the appendices for the full query). The implications of this approach are two-fold:

1. Only content that explicitly referenced one of these London-related terms was included in the dataset. Tweets about social integration issues in London that did not mention one of the London-related terms would have been missed from the data collection.

2. Content was not limited to researcher-defined social integration topics, as would have been the case had we built search queries around keywords related to social integration (e.g. social cohesion, feelings of belonging etc.). This means the conversations captured should give a truer representation of the full spectrum of conversations around social integration.

19-013606-01| Version 1 | Internal and client use only | This work was carried out in accordance with the requirements of the international quality standard for Market Research, ISO 20252, and with the Ipsos MORI Terms and Conditions which can be found at http://www.ipsos-mori.com/terms. © Greater London Authority 2019

There is no objective way of assessing the extent to which these two factors led to the data being representative or unrepresentative. Our judgement, and the rationale for selecting this approach, is that Tweets that mention a London-related term are unlikely to be consistently different – whether in terms of the type of user that posted the Tweet or the content of the Tweet - from those that do not mention a London-related term, at least in terms of the variables that matter for this research (for example, it is not an issue that Tweets mentioning a London-related term are more likely to be made by users who are in London). While it is inevitable that there are some differences between the missing data and the captured data, if this assumption is correct, these differences should be relatively small.

Missing data due to the geo-location filter

A secondary cause of missing relevant data are the filters that are applied before the data is exported from Synthesio. As described in the methodology chapter, filters on both location and language were applied to the data to increase its relevance.

The research question specifically concerned Tweets from users who were residents of London. Based on publicly available data about Twitter users, it is not possible to identify whether a user resides in London. However, in an attempt to limit Twitter users to those present in London, the data was filtered to include only Tweets that originated from London based on the geographical metadata available. When working with Tweets, there are two classes of geographical metadata available:

- Tweet location: available when user shares location at the time of the Tweet.
- Account location: based on the 'home' location provided by the user in their public profile.

Only Tweets where either the Tweet location or account location were stated as London, or an area within London, were included in the data output. Of the Tweets identified by the search query over the search period[27], 51% were geo-located within London. This high proportion is expected given that all the Tweets identified by the search query will have mentioned a London-related terms[28].

Approximately three in ten (29%[29]) Tweets identified by the search query had no geo-location metadata associated with them and were therefore excluded from the dataset. There is no accurate way of assessing the number of these Tweets which originated from London, however it seems likely that Tweets without associated geo-location data are equally likely to originate from London as those with associated geo-location data. This would suggest that approximately 15% of Tweets identified by the search query were excluded from the dataset despite originating in London. Ultimately, this only has an impact on the representativeness of the dataset if those Tweets without associated geo-

[27] The search query itself was only run on data that originated from the UK. Twitter requires all accounts to be associated with the country that the user lives in (for the purposes of customisation and to ensure that Twitter abides by the laws of the relevant country). Therefore, concerns about representatively related to the use of a UK filter within the search query itself are very low.
[28] Of the 2,082,021 mentions identified by the search query, 1,064,812 were geo-located within London.
[29] Of the 2,082,021 mentions identified by the search query, 609,897 had no associated geo-location metadata available.

location metadata are consistently different from those with associated metadata. Again, there is no objective way of assessing this, but it seems likely that any differences would be relatively small.

Missing data due to the language filter

The second filter that was applied within Synthesio was to limit the exported data to English language Tweets only. This relies on Synthesio's custom language identification system which is designed for social media messages and can correctly detect 80 languages, even on very short messages. The reason for limiting the dataset to English language content were practical, as the use of natural language processing in topic modelling requires the dataset to contain a single language.

The vast majority of Tweets (95%[30]) that were geo-located within London were identified as English language. Around 3% of Tweets were unrecognisable as a language (this could be due to one of a number of reasons, for example because the Tweet used a combination of languages, or because the content was limited to a URL or image). Of the 5% non-English language Tweets, the most frequently occurring languages (accounting for more than 1,000 Tweets) are listed in Table 6.4.

Table 1.4: Most frequently excluded non-English languages

Language	Number of Tweets excluded	Excluded Tweets as a proportion of all Tweets geo-located in London
Dutch	2,999	0.3%
Somali	2,247	0.2%
French	2,123	0.2%
Arabic	2,038	0.2%
Italian	2,030	0.2%
Slovak	1,865	0.2%
Spanish	1,488	0.1%
Chinese	1,338	0.1%
Afrikaans	1,098	0.1%

The exclusion of content in languages other than English has implications for the representativeness of the data as these users are likely to be consistently different from users who are posting in English. Additionally, given the focus of this research on social integration, and the additional barriers that non-English language speakers may face integrating, non-English language data may be particularly insightful. However, given the volume of non-English language content identified was so small even if

[30] Of the 1,064,812 Tweets that were geo-located in London, 1,007,143 were identified as English language.

it had been included in the data set, it is unlikely it would have had a large impact on the findings of the analysis (although it may have provided an interesting comparison with English-language data).

In conclusion, the Twitter data is not representative of the population in London; both due to the profile of Twitter users and the way in which Twitter data was extracted for use in this research. Therefore, findings cannot be generalised to the wider population of London. Analysis of Twitter data can provide insight on what Twitter users say about social integration in London but cannot be seen as representative of all Londoners' views. Although not representative however, analysis of social media data can provide a rich understanding on the range of dialog surrounding social integration, leaving us more informed. For example, it can provide access to the opinions and experiences of groups of people (e.g. young men) who are difficult to engage through traditional research approaches, and whose opinions may therefore be underrepresented.

19-013606-01| Version 1 | Internal and client use only | This work was carried out in accordance with the requirements of the international quality standard for Market Research, ISO 20252, and with the Ipsos MORI Terms and Conditions which can be found at http://www.ipsos-mori.com/terms. © Greater London Authority 2019

Learnings and future considerations

One of the objectives of this research has been to consider the extent to which the analysis of social media data – specifically Twitter data - can be used by the GLA to monitor social integration across London. To this end, this chapter outlines several observations the research team have made to both the findings and the process of social media data for this project, highlighting the challenges and opportunities for this kind of approach.

1. Data relating to social integration is rich in depth and breadth

Despite the bottom-up approach that was taken in the early stages of the analysis – with no specific focus on social integration – relevant issues were salient in the themes identified by the topic model. This confirms that, as suggested by the GLA's scoping study, Twitter is a rich source of information relating to social integration in London.

Furthermore, although the analysis has not identified any completely new areas of social integration that are not accounted for in the GLA's existing framework, the range of topics discussed by Twitter users is very varied (except for topics related to relationships – as described later in this chapter). The 'bottom-up' research methodology has presented a broader picture of social integration than could traditionally be captured through survey research (for example through a pre-coded question). Moreover, the data contains additional anecdotal information about the impact of social integration – both positive and negative. For example, the practical impacts of being unable to access transport, the financial impact of regeneration on small businesses, the positive impact that volunteers have on the communities they serve.

This provides useful insight into the real-life day-to-day experiences of social integration. This could be used to both inform the way in which the GLA measures social integration across London, and to aid understanding of the consequences of poor and successful social integration more generally.

2. Some areas of social integration are more present in the data than others

The bottom-up topic modelling succeeded in identifying many of the key elements of social integration within the cleaned data corpus. Notably, there were a substantial number of topics relating to participation. However, some topics of interest were missing from the topic model; particularly those relating to relationships. For example, topics relating to social mixing between people from different backgrounds and helping neighbours did not emerge from the topic modelling.

It is possible that, compared with topics relating to participation, topics relating to relationships are less frequently discussed on Twitter due to their more personal nature. Indeed, our top-down analysis of the data corpus using additional search queries (as detailed in the 'top-down analysis' chapter) failed to identify significant conversation on these topics.

Alternatively, it is possible that the search query that was used to collect data from the Twitter corpus excluded data pertaining to these topics (and that they are therefore not represented in the data corpus that was analysed). One facet of the approach which may have caused this was the

19-013606-01| Version 1 | Internal and client use only | This work was carried out in accordance with the requirements of the international quality standard for Market Research, ISO 20252, and with the Ipsos MORI Terms and Conditions which can be found at http://www.ipsos-mori.com/terms. © Greater London Authority 2019

requirement that each Tweet in the corpus contains a London-related term. It may be the case that Tweets discussing topics relating to relationships tend not to include London-related terms and are therefore excluded from the data corpus.

At this exploratory stage of analysis, it was necessary to specify the inclusion of a London-related term within the Tweet in order to ensure that the content collected was related to London (as simply being sent from London itself provides no guarantee of this). However, future analysis could remove this requirement - at least in any data collection that aims to collect data pertaining to relationships - to increase the likelihood of these topics being represented in the data.

3. Further research could explore approaches to distinguishing between individuals and organisations

The research aimed to analyse Tweets sent by those living in London and about London specific topics and issues. While the search query succeeded in collecting data almost exclusively about London specific topics and issues, the Tweets were sent from a combination of both individuals and organisations. There would possibly be additional insight to be gained from analysing Tweets sent by individuals separately from those sent by organisations.

However, distinguishing between organisations and individuals is challenging for two reasons. On a practical level, Twitter provides no metadata to indicate whether an account is held by a private individual or an organisation, so it is not possible to filter the data collected on this variable. On another level, there are challenges in defining 'organisations' and 'individuals', with it being unclear how some accounts (for example accounts belonging to informal neighbourhood groups, or small cycling groups) should be defined.

4. Future waves of research could take a more 'top-down' approach to deliver more comparable analysis

This project has taken a bottom-up approach to data collection and analysis. Collecting the widest possible definition of data, and basing analysis of the content included within these posts. This is a crucial stage in understanding the potential value of this social media data. However, future waves could benefit from a different approach, which would build upon the bottom-up analysis to specify top-down search queries, based on the key issues of concern. For example, building a query to identify a specific social integration topic. Using this top-down approach, coupled with geo-tagging in London, could create results that would be comparable across waves. However, it would also be important to continually review the search queries to make sure that they maintained relevance, and captured the most recent topics of conversation.

5. Further research could assess the extent to which social media data could be used as an early warning system for emerging social integration issues

Mapping the data collected for this project, or future waves of research (potentially collected in a top-down manner) alongside other data sources collected could help assess whether Londoners were raising issues on social media prior to identification of key issues on other primary research or

19-013606-01| Version 1 | Internal and client use only | This work was carried out in accordance with the requirements of the international quality standard for Market Research, ISO 20252, and with the Ipsos MORI Terms and Conditions which can be found at http://www.ipsos-mori.com/terms. © Greater London Authority 2019

secondary data. Without further triangulation of data, it is unclear whether the posts on social media could be used as an indication of emerging social integration concerns or harms.

19-013606-01| Version 1 | Internal and client use only | This work was carried out in accordance with the requirements of the international quality standard for Market Research, ISO 20252, and with the Ipsos MORI Terms and Conditions which can be found at http://www.ipsos-mori.com/terms. © Greater London Authority 2019

Appendices

1.16Final search query

(bexley OR (brent NOT (oil OR crude OR usd OR gbp OR shell OR david OR faiyaz)) OR "city of london" OR haringey OR havering OR hillingdon OR islington OR (kingston NOT (jamaica OR town)) OR lambeth OR newham OR redbridge OR southwark OR sutton OR "tower hamlets" OR "waltham forest" OR acton OR "angel edmonton" OR "bakers arms" OR bankside OR "barking riverside" OR "borough high street" OR "borough market" OR "brent street" OR "brick lane" OR "tottenham high road" OR camberwell OR "canada water" OR "cheam village" OR cheapside OR "chrisp street" OR "collier row" OR "the strand" OR "crouch end" OR heathway OR downham OR "lordship lane" OR "earls court" OR "east beckton" OR greenwich OR "edgware road" OR (elephant NEAR/1 castle) OR euston OR feltham OR "fleet street" OR "fulham road" OR "green lanes" OR "green street" OR "upton park" OR "harold hill" OR kingsway OR "king's road" OR "kings road" OR "lavender hill" OR "queenstown road" OR "leadenhall market" OR "lee green" OR "lower marsh" OR ("the cut" AND london) OR "mare street" OR marylebone OR "merrielands crescent" OR "merry fiddlers" OR "muswell hill" OR "new addington" OR cheam OR chingford OR "portobello road" OR "praed street" OR paddington OR "queensway" OR "westbourne grove" OR rosehill OR selsdon OR "shepherd's bush" OR norwood OR "st john's wood" OR "st johns wood" OR "temple fortune" OR "upper norwood" OR "crystal palace" OR "victoria street" OR "walworth road" OR "warwick way" OR "tachbrook street" OR "watney market" OR "wentworth street" OR "west end" OR "west green road" OR "seven sisters" OR (westfield NEAR/2 london) OR (westfield NEAR/2 stratford) OR (westfield NEAR/2 "white city") OR (westfield NEAR/2 "shepherd's bush") OR whetstone OR yiewsley OR "west drayton" OR (barking NOT (dog OR animal OR vet OR mad)) OR battersea OR (bow NOT ("take a bow" OR "bow down")) OR "brent north" OR isleworth OR wallington OR fulham OR "woodford green" OR "west norwood" OR ealing OR "east ham" OR edmonton OR eltham OR thamesmead OR (heston NOT (blumenthal OR dinner OR duck)) OR "golders green" OR "stoke newington" OR shoreditch OR hammersmith OR harrow OR harlington OR hendon OR upminster OR ilford OR finsbury OR surbiton OR lewisham OR penge OR wanstead OR morden OR sidcup OR orpington OR (poplar NOT (tree OR wood)) OR limehouse OR richmond OR romford OR pinner OR (twickenham NOT rugby) OR (vauxhall NOT (@vauxhall OR astra OR corsa OR chevette OR car OR zafira OR vectra OR insignia OR adam OR grandland OR crossland OR mokka OR viva)) OR westminster OR "abbey wood" OR "albany park" OR "alexandra palace" OR anerley OR "angel road" OR barnehurst OR "barnes bridge" OR bellingham OR belmont OR belvedere OR berrylands OR bexleyheath OR bickley OR birkbeck OR blackheath OR "bowes park" OR brentford OR brimsdown OR brockley OR bromley OR brondesbury OR "brondesbury park" OR "bruce grove" OR "bush hill park" OR barnsbury OR "cambridge heath" OR camden OR canonbury OR carshalton OR "castle bar park" OR catford OR "chadwell heath" OR charlton OR chelsfield OR chessington OR chislehurst OR chiswick OR clapton OR coulsdon OR crayford OR "crews hill" OR cricklewood OR "crofton park" OR "crouch hill" OR dagenham OR dalston OR kingsland OR "denmark hill" OR deptford OR "drayton green" OR "drayton park" OR earlsfield OR croydon OR dulwich OR ("eden park" AND London) OR "edmonton green" OR "elmers end" OR "elmstead woods" OR "emerson park" OR enfield OR erith OR "essex road" OR falconwood OR "fenchurch street" OR frognal OR "forest gate" OR "forest hill" OR fulwell OR "gidea park" OR "gipsy

hill" OR goodmayes OR "gordon hill" OR "gospel oak" OR "grange park" OR "grove park" OR hackbridge OR hackney OR "hadley wood" OR haggerston OR hampton OR "hampton wick" OR hanwell OR "harold wood" OR harringay OR "hatch end" OR "haydons road" OR (hayes AND london) OR "harlington" OR "headstone lane" OR "herne hill" OR "highams park" OR "hither green" OR homerton OR "honor oak park" OR hornsey OR hoxton OR "imperial wharf" OR kenley OR "kent house" OR "kentish town" OR kew OR kidbrooke OR kilburn OR knockholt OR ladywell OR "lea bridge" OR "midland road" OR limehouse OR "london fields" OR "loughborough junction" OR "malden manor" OR "manor park" OR "maze hill" OR "mill hill" OR mitcham OR "morden south" OR mortlake OR "motspur park" OR mottingham OR (barnet NOT hair) OR beckenham OR "new cross" OR "new eltham" OR "new malden" OR "new southgate" OR norbiton OR norbury OR "northolt park" OR "northumberland park" OR "norwood junction" OR nunhead OR "oakleigh park" OR "palmers green" OR peckham OR "petts wood" OR plumstead OR "ponders end" OR purley OR "purley oaks" OR "queens road" OR "peckham" OR "queenstown road" OR rainham OR ravensbourne OR "raynes park" OR "rectory road" OR reedham OR riddlesdown OR rotherhithe OR sanderstead OR selhurst OR "seven kings" OR shadwell OR shortlands OR sidcup OR "silver street" OR "slade green" OR bermondsey OR greenford OR hampstead OR (merton NOT (college OR paul)) OR "south tottenham" OR (southall NOT neville) OR southbury OR "st helier" OR "st james street" OR "st johns" OR "st margarets" OR "st mary cray" OR "st pancras" OR "stamford hill" OR (stratford NOT (avon OR shakespeare)) OR "strawberry hill" OR streatham OR sudbury OR "sundridge park" OR "surrey quays" OR "sutton common" OR sydenham OR "syon lane" OR "teddington" OR "thornton heath" OR tolworth OR "tulse hill" OR "turkey street" OR "upper holloway" OR waddon OR wandsworth OR "wanstead park" OR wapping OR "west sutton" OR "west wickham" OR "westcombe park" OR "white hart lane" OR whitton OR "winchmore hill" OR "wood street"OR "woodgrange park" OR woodmansterne OR woolwich OR "worcester park" OR "bushy park" OR "regent's park" OR "regents park" OR "regent park" OR "hyde park" OR "st james park" OR "victoria park" OR "crystal palace park" OR "alexandra park" OR "brockwell park" OR "thames chase" OR "epping forest" OR "trent park" OR "hainault forest country park" OR "wormwood scrubs" OR "postman's park" OR "ally pally" OR "lee valley" OR "clissold park" OR "holland park" OR "olympic park" OR "springfield park" OR "burgess park" OR "painshill park" OR "grosvenor square" OR "gunnersbury park" OR "bishop park" OR "osterley park" OR "pymmes park" OR "oak hill park" OR "wanstead flats" OR "canon's park" OR "crane park" OR "foots cray meadows" OR "gladstone park" OR "grovelands park" OR "hilly fields" OR "northala fields" OR "hoblingwell" OR "old deer park" OR "orpington park" OR "riddleshaw park" OR "roundshaw downs" OR "south norwood park" OR "valentines park" OR "victoria dock" OR aldgate OR alperton OR (angel AND london) OR archway OR "arnos grove" OR ("baker street" NOT (rafferty OR sherlock OR holmes OR song)) OR balham OR barbican OR barkingside OR "barons court" OR bayswater OR becontree OR "belsize park" OR "bethnal green" OR blackfriars OR "blackhorse road" OR "bond street" OR "boston manor" OR "bounds green" OR "bow road" OR brixton OR bromley-by-bow OR "burnt oak" OR "caledonian road" OR "canary wharf" OR "canning town" OR "cannon street" OR "canons park" OR "chalk farm" OR "chancery lane" OR "charing cross" OR "chiswick park" OR clapham OR cockfosters OR colindale OR "colliers wood" OR "covent garden" OR "dollis hill" OR "earl's court" OR finchley OR eastcote OR edgware OR "elm park" OR "victoria embankment" OR "albert embankment" OR fairlop OR farringdon OR "finsbury park" OR "fulham roadway" OR "gants hill" OR "gloucester road" OR "goldhawk road" OR "goodge street" OR "grange hill" OR "great

portland street" OR "green park" OR gunnersbury OR hainault OR "hanger lane" OR harlesden OR wealdstone OR "harrow-on-the-hill" OR "hatton cross" OR "hendon central" OR kensington OR "highbury" OR "islington" OR highgate OR holborn OR "holland park" OR "holloway road" OR "hornchurch" OR hounslow OR "hyde park corner" OR ickenham OR kennington OR kensal OR "king's cross" OR "st pancras" OR kingsbury OR knightsbridge OR "ladbroke grove" OR "lancaster gate" OR "latimer road" OR "leicester square" OR (leyton NOT dunlop) OR leytonstone OR "liverpool street" OR "london bridge" OR "maida vale" OR ("manor house" AND london) OR ("mansion house" AND london) OR "marble arch" OR "mile end" OR moorgate OR "mornington crescent" OR neasden OR "newbury park" OR wembley OR northfields OR northolt OR "northwick park" OR northwood OR "northwood hills" OR "notting hill" OR oakwood OR "old street" OR osterley OR ((oval AND london) NOT (kia OR cricket))OR "oxford circus" OR paddington OR "park royal" OR "parsons green" OR perivale OR "piccadilly circus" OR pimlico OR plaistow OR "preston road" OR putney OR (("queen's park" OR "queens park") NOT "park rangers") OR queensbury OR queensway OR "ravenscourt park" OR "rayners lane" OR "royal oak" OR "russell square" OR "seven sisters" OR "sloane square" OR snaresbrook OR kenton OR (wimbledon NOT (murray OR federer OR tennis OR match OR game OR "#wearewimbledon" OR "#afcwimbledon")) OR woodford OR southfields OR (southgate NOT gareth) OR "st james's park" OR "st james' park" OR "st paul's" OR "st pauls" OR "stamford brook" OR stanmore OR "stepney green" OR stockwell OR "stonebridge park" OR "swiss cottage" OR (temple AND london) OR tooting OR "tottenham court road" OR "tottenham hale" OR "totteridge" OR "whetstone" OR "tower hill" OR "tufnell park" OR "turnham green" OR "turnpike lane" OR "upminster bridge" OR upney OR uxbridge OR walthamstow OR "warren street" OR ("warwick avenue" NOT (song OR duffy)) OR (waterloo NOT (sunset OR kinks OR song)) OR "west brompton" OR ruislip OR "westbourne park" OR "white city" OR whitechapel OR willesden OR "wood green" OR "wood lane" OR "woodside park" OR "abbey road" OR "beckton" OR "blackwall" OR "cutty sark" OR "devons road" OR "elverson road" OR "gallions reach" OR "heron quays" OR "island gardens" OR "langdon park" OR "mudchute" OR "pudding mill" OR "south quay" OR "star lane" OR "tower gateway" OR "west silvertown" OR "westferry" OR "addington village" OR "addiscombe" OR "ampere way" OR "beddington lane" OR "belgrave walk" OR "blackhorse lane" OR "dundonald road" OR "coombe lane" OR "elmers end" OR "fieldway" OR "dundonald road" OR "harrington road" OR "king henry's drive" OR "lebanon road" OR "lloyd park" OR "merton park" OR "new addington" OR "phipps bridge" OR "reeves corner" OR "sandilands" OR "therapla lane" OR "crystal palace" OR "waddon march" OR "wandle park" OR "wellesley road" OR "south london" OR "north london" OR "west london" OR "east london" OR "south east london" OR "north east london" OR "south west london" OR "north west london" OR "se london" OR "nw london" OR "ne london" OR "sw london") NOT(((delay OR traffic OR accident) AND (tfl OR junction OR tube)) OR salary OR job* OR staff OR "guaranteed hours" OR vacanc* OR part-time OR "part time" OR full-time OR "full Time" OR staff* OR "flat to rent" OR "house to rent" OR "double room" OR "double bedroom" OR "single room" OR "single bedroom" OR "bedroom flat" OR "bedroom flat" OR "bedroom house" OR (sale AND (property OR house OR bedroom OR flat OR apartment)) OR (rent AND (property OR house OR flat OR apartment)) OR coupon OR coupons OR couponing OR cupon OR cupons OR cuponing OR voucher OR vouchers OR sale OR sales OR deal OR deals OR rebate OR offer OR ebay OR hotdeals OR dailydeals OR dailydeal OR discount OR hotdeal OR "free ship" OR "free shipping" OR freeship OR freeshipping OR

"cash back" OR cashback OR giveaway OR giveaways OR freebie OR freebies OR freebiefriday OR promo OR save OR "special offer" OR ticket* OR twickets)

1.17Topic modelling process

Various different models were tested for the topic modelling, and Non-Negative Matrix Factorisation (NMF) was chosen to generate 80 topics. With more than 80, the additional topics increasingly seemed to be similar to existing ones (several about different aspects of healthcare for example, or sunset/sky, skyline/photography, capture/focus/picture all coming through separately) and even with 80 there are still a few that were grouped when manually reviewed. With fewer than 80, there were some potentially interesting topics (such as museums or social isolation) that didn't come out.

NMF is a form of matrix factorisation where a recursive algorithm generates a user-specified number of other features that, in combination, approximate the initial matrix as accurately as possible. The usage in topic modelling is to decompose a term-document matrix where the features created are based on word co-occurrence and so imbue a kind of context. As an example 'running' might frequently occur with 'marathon' or 'fitness' and so contribute towards an exercise topic.

Once these features have been modelled we can calculate a component score for each sentence in our datafile by topic. The greater the score, the more terms and higher weighted terms there were for that topic in that sentence, a low score would indicate very little relevance to that topic.

With NMF, initially every post is given a relevance score in every topic. Having identified the potentially interesting topics (excluding for example discussion around dates, ages, weblinks) we manually determined the cut off for each topic where we felt there was still a clear theme coming through and excluded posts with relevance below that. These cleaned topics were then assessed manually for relevance to the project goals.

After this, to see if there were further topics we could bring out that were otherwise being obscured by the larger repetitions, we tried running an additional set of topics only using those posts not currently in any other topics. This generated 50 new topics, and of those, only three were deemed relevant to social integration.

19-013606-01| Version 1 | Internal and client use only | This work was carried out in accordance with the requirements of the international quality standard for Market Research, ISO 20252, and with the Ipsos MORI Terms and Conditions which can be found at http://www.ipsos-mori.com/terms. © Greater London Authority 2019

1.18Top-down search queries

Topic	Full query
School activities and events	school/LEMMA
Local area change and affordability	gentrification/LEMMA, gentrify/LEMMA, gentrified/LEMMA, close/LEMMA 2/dist down/LEMMA, die|dying|drive|driven/LEMMA 2/dist out/LEMMA, new|house|housing/LEMMA 2/dist develop|development|build/LEMMA, new/LEMMA 2/dist build/LEMMA, luxury|penthouse/LEMMA 3/dist2 flat|flats|apartment|appartments|develop|development|housing/LEMMA, rent/LEMMA, landlord/LEMMA, price/LEMMA 2/dist out/LEMMA, unaffordable/LEMMA, house|apartment|flat|property|home/LEMMA 5/dist2 price|prices/LEMMA, cheap|budget|affordable/ LEMMA 6/dist2 housing|home|homes|flat|flats|accommodation/ LEMMA, evict|evicted|eviction/LEMMA],
Conversations with friends	friend|mate|pal|buddy/LEMMA, friendship/LEMMA
Helping neighbours	favour|lend|borrow/LEMMA, neighbourhood/LEMMA, local/LEMMA 4/dist2 resident|residents|men|women|people|community/LEMMA

19-013606-01| Version 1 | Internal and client use only | This work was carried out in accordance with the requirements of the international quality standard for Market Research, ISO 20252, and with the Ipsos MORI Terms and Conditions which can be found at http://www.ipsos-mori.com/terms. © Greater London Authority 2019

Steven Ginnis
Research Director
steven.ginnis@ipsos.com

Sylvie Hobden
Associate Director
sylvie.hobden@ipsos.com

Pete Duncan
Associate Director
pete.duncan@ipsos.com

Emily Mason
Research Executive
emily.mason@ipsos.com

Imogen Drew
Research Executive
imogen.drew@ipsos.com

For more information

3 Thomas More Square
London
E1W 1YW

t: +44 (0)20 3059 5000

www.ipsos-mori.com
http://Twitter.com/IpsosMORI

About Ipsos MORI's Social Research Institute

The Social Research Institute works closely with national governments, local public services and the not-for-profit sector. Its c.200 research staff focus on public service and policy issues. Each has expertise in a particular part of the public sector, ensuring we have a detailed understanding of specific sectors and policy challenges. This, combined with our methods and communications expertise, helps ensure that our research makes a difference for decision makers and communities.