George Downing
J7526527

Digital Inequalities within the United Kingdom

Executive Summary

Internet and mobile access play a vital part in everyday life, both for consumers and business, with studies relating greater access to economic growth and increased productivity.

This report will investigate UK broadband and mobile network coverage, to highlight inequalities. It will answer two main questions:

1. Has UK mobile network coverage improved over the last five years?
2. Is there a correlation between poor broadband coverage, and poor mobile network coverage?

The report is prepared for the board of a non-commercial organisation focused on reducing UK digital inequality.

The investigation used two datasets from Ofcom with information on broadband and mobile network coverage across UK local authorities between 2019 and 2023. The datasets were combined to assess the relationship between key variables. The datasets do not include any person-identifying data, so there are no deanonymisation risks.

The investigation revealed that local authority regions with no providers of reliable 4G signal has steadily decreased over the last 5 years, with significant outliers reducing. Success has occurred in the southwest of England and Wales, however, there has been a slight rise in areas surrounding London.

Cluster analysis was performed and revealed two clear clusters. One of regions with a high percentage of premises with poor broadband and mobile network coverage, and another with a low percentage of such premises. Roughly 50 areas fell into the former cluster where further investment is still required.

Testing suggested a strong positive linear relationship between poor mobile network and broadband coverage (Pearson's R test R-value of 0.764 with 1 indicating a total positive correlation) and that the results were statistically significant (p-value of 0.0 indicating strong evidence against no correlation).

There are limitations to these results, including potential overlap between the clusters as well as removal of data for outdated local authorities resulting in incomplete data. As regards the latter, the dataset remained sufficient to provide reliable results. Furthermore, the report does not identify which specific areas have low coverage. Researching this would establish a regional priority action list.

This investigation suggests it is vital to target areas with poor coverage for improvement, as they are likely to suffer from both poor mobile and broadband coverage, which can greatly impact work and personal life.

Further research is recommended to investigate the underlying reasons behind the trends to improve the impact of investment.

(389 words)

Aims and Objectives

This report investigates the quality of broadband and mobile network coverage in the United Kingdom to investigate cases of digital inequality within the UK. It will use a dataset detailing the mobile network coverage ('mobile coverage') from different providers for premises in different areas, combined with a broadband coverage dataset for different local authorities.

Two key questions will be considered; has the quality of mobile coverage improved within the UK over the last five years, and is there a correlation between areas with poor mobile coverage and areas with poor broadband coverage? The analysis from these questions will give insight into whether the current methods addressing digital inequalities are effective, and highlight which regions need the investment.

Background

The availability of broadband and mobile coverage has become increasingly important for consumers and businesses. Studies have shown broadband connection "enhances quality-of-life for households and facilitates economic and social opportunities" (Valentín-Sívico *et al*, 2023), and mobile coverage increases productivity and reduces time loss (Building Digital UK, 2022).

The report is prepared for the board of a non-commercial organisation focused on reducing UK digital inequality.

The investigation combines the datasets of mobile coverage and broadband coverage to view relationships between these two issues. The broadband dataset was previously investigated (Downing, 2024). This dataset combination will show the relationship between the issues and identify whether there is an underlying trend of regions lacking coverage.

The datasets contain information on coverage, however, no data on potential causal factors for inequalities. This limitation means the analysis can be used to identify underperforming areas but cannot offer insight into underlying causes.

This report uses 4G mobile coverage as an indicator of good mobile coverage, being the most common cellular network, with users being connected roughly 88% of the time (Ofcom, 2023a). Indoor and outdoor premises (‘premises’) with service from no providers will be classified as poor coverage and those with three or four providers as good coverage.

The Universal Service Obligation (‘USO’) will be used as a baseline for broadband coverage which is the standard set by Ofcom, requiring “decent and affordable broadband connection” (Ofcom, 2023b).

Sources of Data

Both datasets are available from Ofcom’s ‘Connected Nations’ reports (Ofcom, 2023c) under the Open Government Licence v3.0 which states that anyone can “copy, publish, distribute and transmit the Information; adapt the Information; exploit the Information” provided they “acknowledge the source of the Information” and “where possible, provide a link to this licence” (The National Archives). All work in this report falls under this licence and is permissible.

The broadband coverage dataset was obtained, imported and cleaned using pandas and finally stored into a document database within MongoDB. The dataset covers all UK local authorities with information on different broadband availability between 2019 and 2023. The dataset’s preprocessing was reviewed previously (Downing, 2024).

The mobile coverage dataset was downloaded from Ofcom and details coverage across all local authorities annually between 2019 and 2023. The data shows the coverage from the 4 main operators (EE, Virgin Media O2, Three and Vodafone) across different premises. All years contain the percentage of premises that met a minimum threshold for different data types, including 4G, and how many providers satisfied this threshold.

There are no concerns surrounding person-identifying data within the datasets as they are focused on premises as a whole and contain no precise location data that would reveal this.

Using this data could have a beneficial impact by identifying underserved areas and allowing for smarter resource allocation into these areas. There are concerns that increased availability could increase property prices, negatively impacting homebuyers, however, this data could also improve freedoms such as working from home.

Combining the datasets enhances investigation into the relationship between broadband and mobile coverage. This could identify whether underlying reasons for poor coverage exist by showing a clear correlation, or if poor coverage is seemingly random. The combination will also allow for clear identification of improved coverage need if the number of areas with poor access to broadband and mobile networks are decreasing.

Analysis Pipeline

The mobile coverage datasets were imported into separated dataframes by year of data collection. Each dataframe was cleaned by checking for outdated local authorities and removing the related rows. A column was added to each dataframe indicating year of data collection. The dataframes were merged into one larger dataframe using an outer join to ensure no data loss. A smaller dataset was created using the relevant columns for this investigation. The dataset was stored in a document database using MongoDB and the null values were set to 0.

To combine the two datasets, the broadband coverage dataset was accessed via MongoDB, and a small dataframe was created with relevant columns. Using the smaller mobile dataframe, averages were calculated for premises with reliable and unreliable 4G signal. These dataframes were then combined on the local authority code and year of collection.

The choice to remove outdated local authorities results in missing data which could impact the results' reliability. However, only two local authorities required removal and therefore the sample will be sufficiently large for analysis. Additionally, some new columns of data were added between each dataset, but as these are unrelated to the specific areas of interest for this investigation, results should not be affected.

To investigate the relationship between mobile and broadband coverage, cluster analysis was performed to establish patterns in the data, alongside changes over time. A k-means clustering algorithm was applied to a dataset of the percentage of premises with no 4G signal, and the percentage of premises below the USO in 2019, 2023 and the full dataset. To do this the combined dataframe was filtered by year, and a smaller one was created showing only these two columns. The algorithm was trained for 2, 3 and 4 clusters for the years stated, and silhouette plots were used to decide the number of clusters that produced optimal results.

In all numbers of clusters, the largest cluster has a higher cluster cohesion than the other clusters in the plot. For figure 1, cluster 0 is much narrower and has weaker cluster cohesion, with fewer values close to 1, as well as some negative values suggesting some cluster overlap. Figures 2 and 3 show similar trends for the narrower clusters but appear to include more negative values in total. Based on the plots, the number of points with a negative cluster coefficient seems to increase as the number of clusters increases suggesting a smaller number of clusters is more appropriate for this dataset. Based on the silhouette plots and the principle of parsimony, two clusters were chosen to be the most appropriate number.

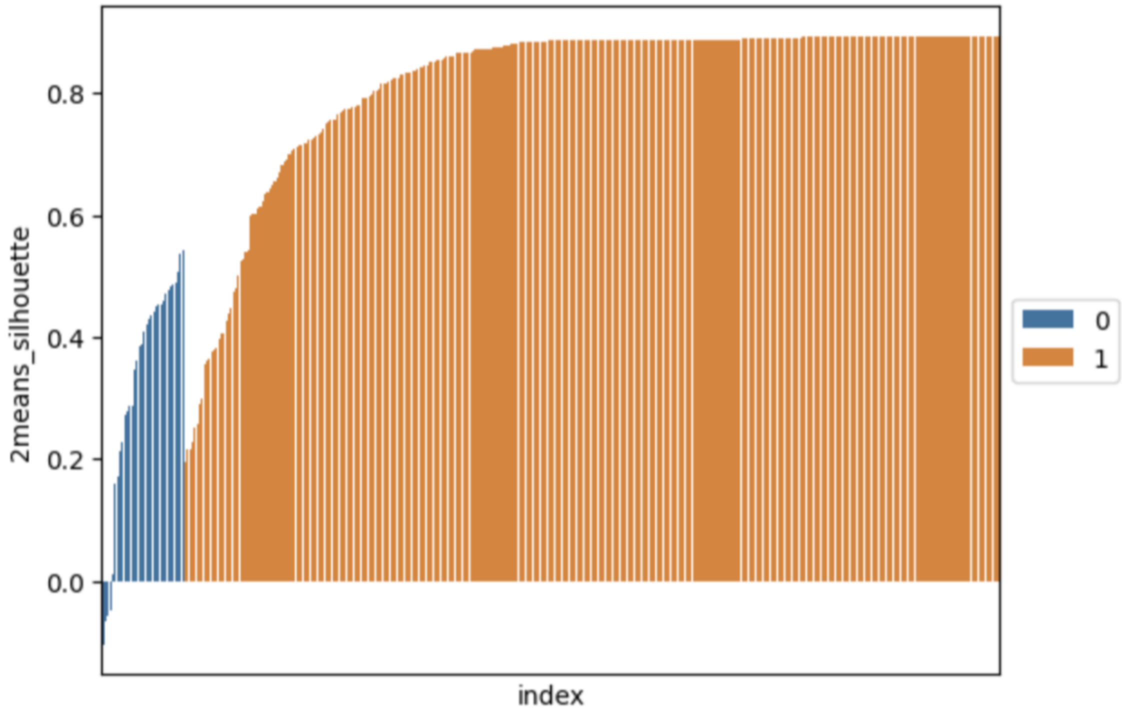

Figure 1: Silhouette plot for 2 means clustering, 2019 dataset

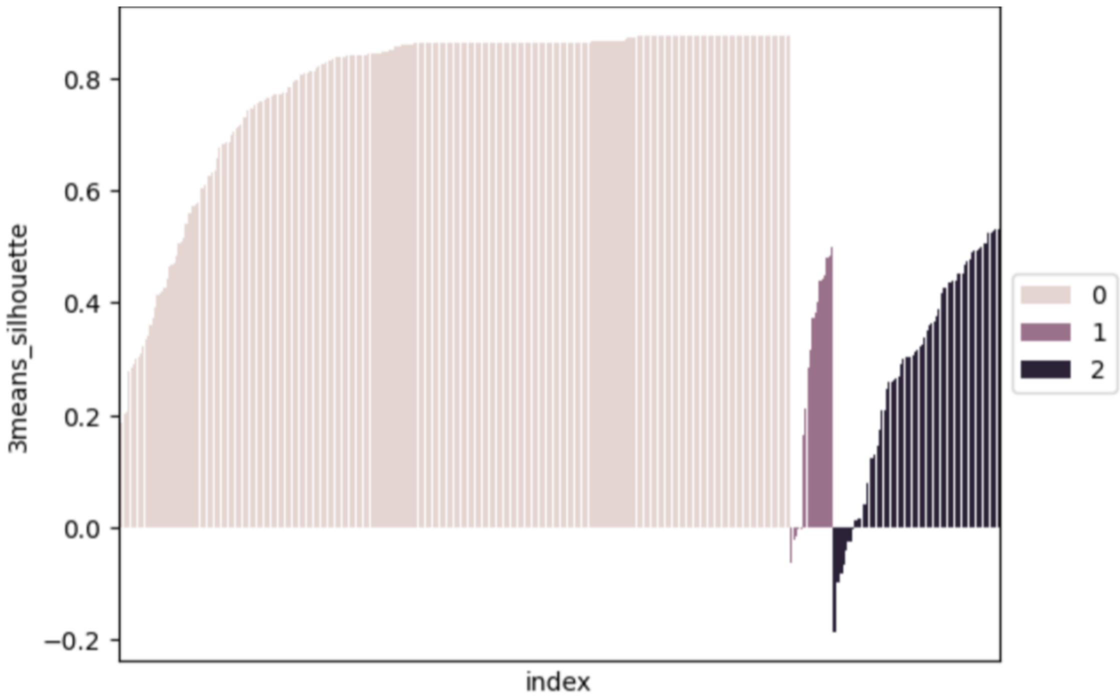

Figure 2: Silhouette plot for 3 means clustering, 2019 dataset

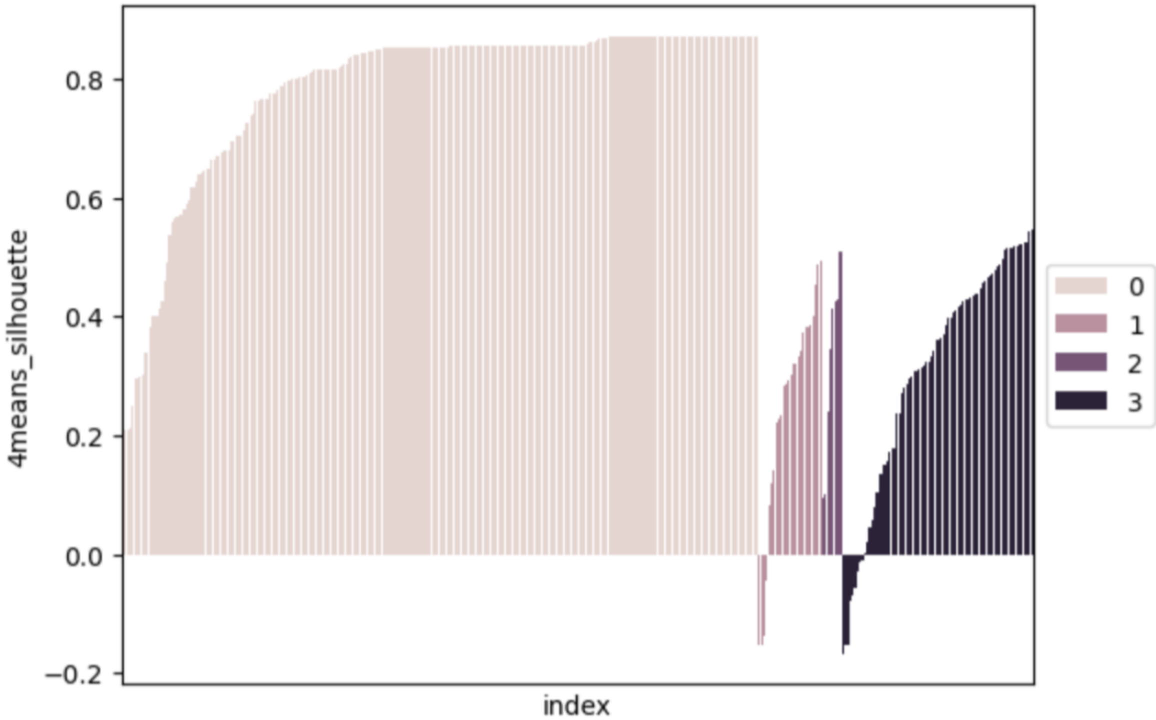

Figure 3: Silhouette plot for 4 means clustering, 2019 dataset

A 2-means clustering algorithm was then trained on the 2023 dataset and the full dataset. The silhouette plots for these clusters are shown in figures 4 and 5. Since these plots showed very similar trends to each other, and the 2019 dataset, it was decided that two clusters was reasonable for showing trends in this data.

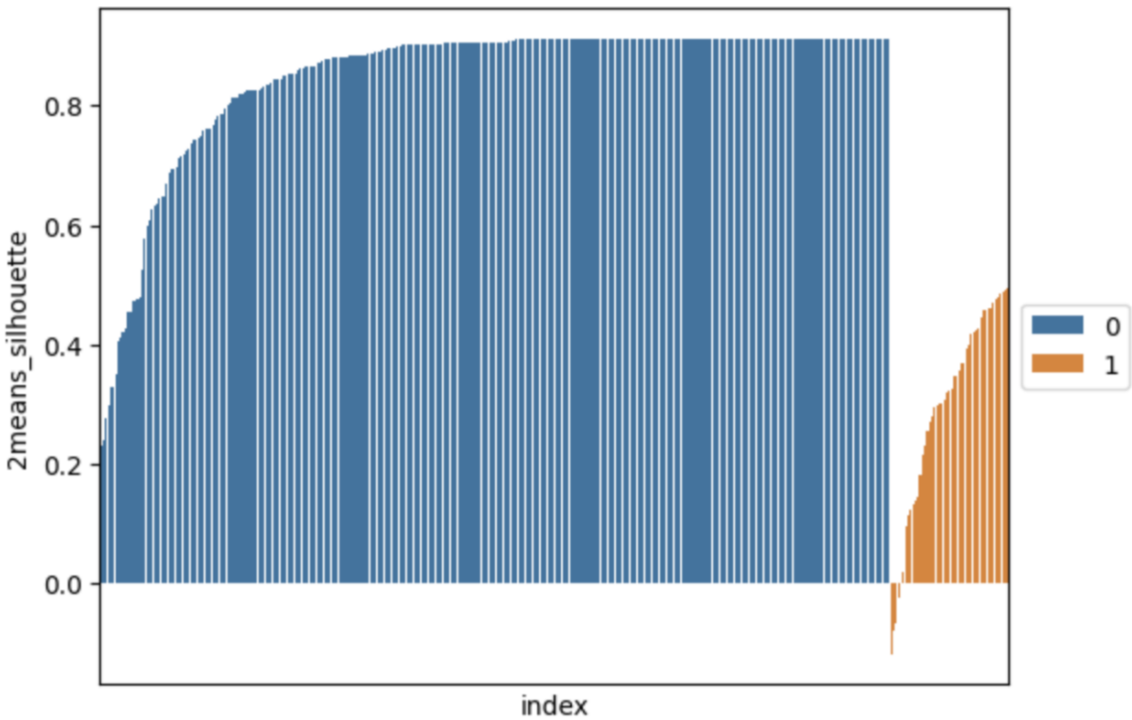

Figure 4: Silhouette plot for 2 means clustering, 2023 dataset

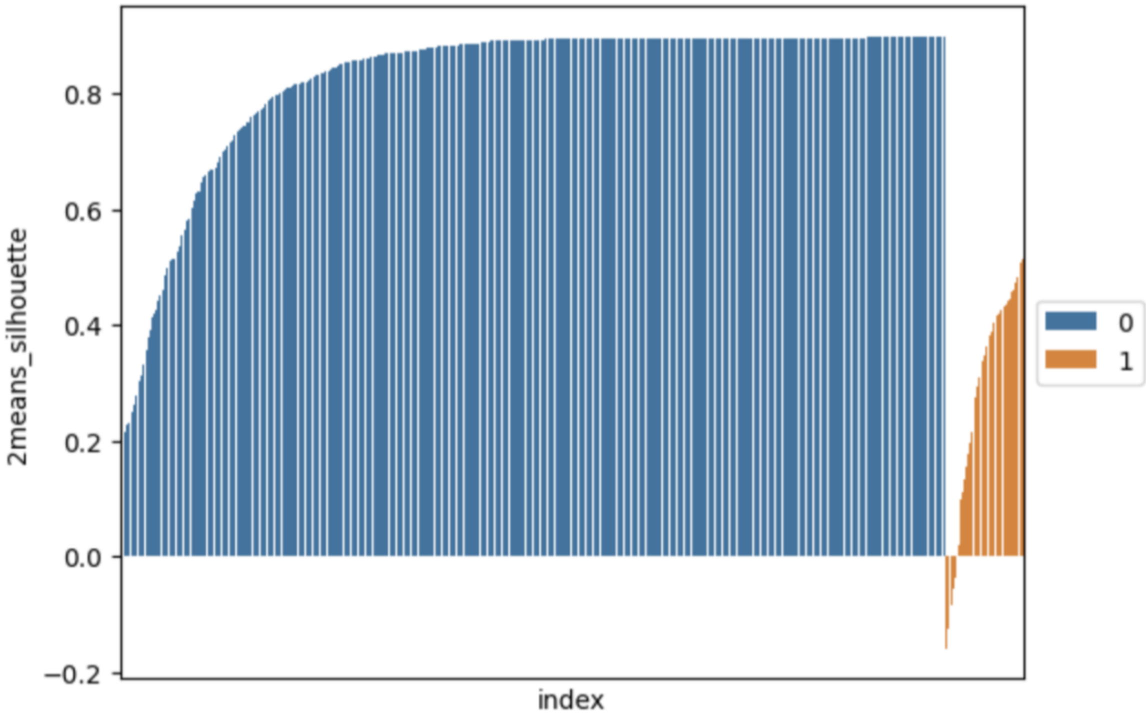

Figure 5: Silhouette plot for 2 means clustering, full dataset

There are limitations to cluster analysis, specifically found in datasets that have non-spherical clusters or those with more outliers, as these can distort the suggested clusters. However, the data seems to fit a spherical cluster reasonably well, and, whilst the data is spread out, the number of outliers is sufficiently small to not be problematic.

Finally, a Pearson's R test was performed to investigate the strength of the relationship between these variables, and whether results were statistically significant. This test is limited in that it only shows linear relationships, although, based on figure 6 the relationship between the two variables investigated appears to be linear.

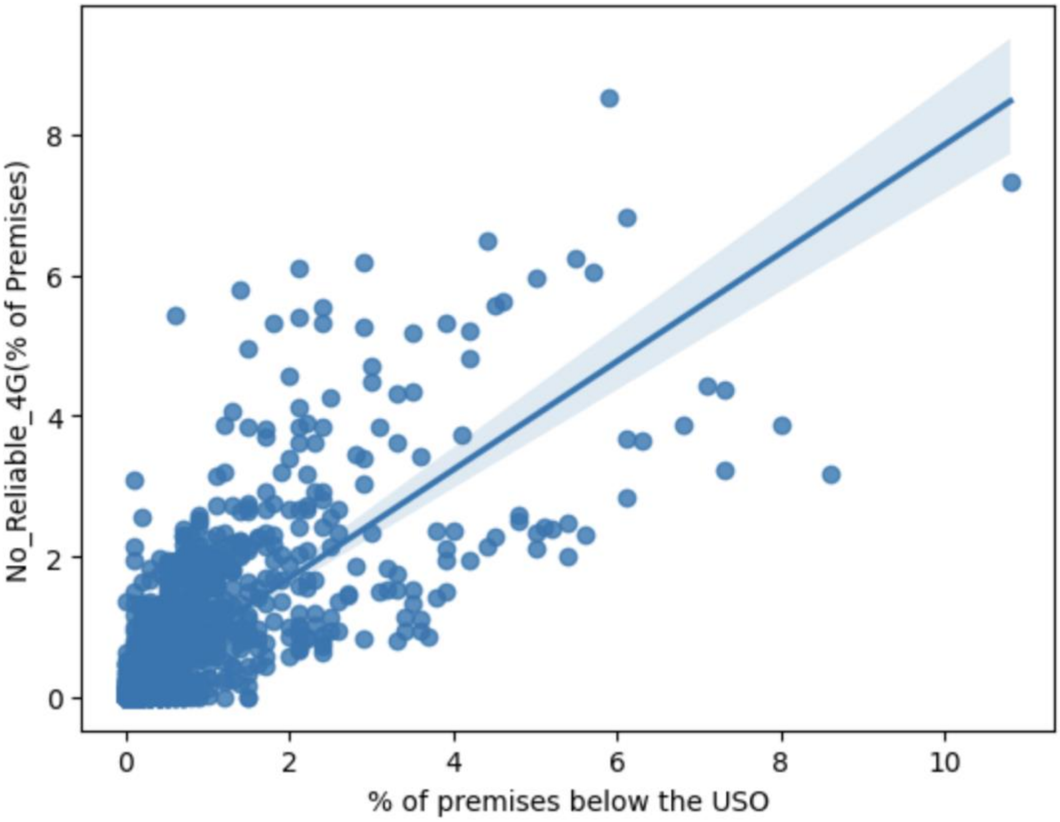

Figure 6: Percentage of premises below the USO vs percentage of premises with no reliable 4G signal

Findings

To investigate whether mobile coverage has improved in the UK over the last 5 years, the average percentage of premises with no reliable 4G signal was selected. Figure 7 shows that there has been slow but gradual improvement over the time frame. The upper quartile has reduced steadily since 2019 and the number and size of high outliers appears to be decreasing. This suggests improved mobile coverage with reduced and less severe digital inequalities. However, the changes in the median and upper quartile are marginal. With the small scale, drawing precise conclusions is difficult, presenting an issue with their use.

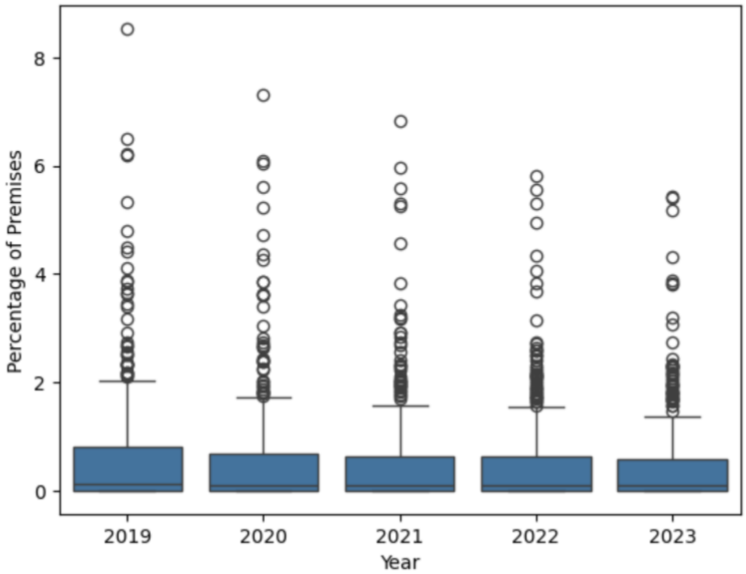

Figure 7: Comparative boxplot of average percentage of premises with no reliable 4G signal by year

Figure 8 investigates the potential for a geographic component in the rate of improvement of mobile coverage, whilst also offering an alternative visualisation for improvement. Figure 8 appears to show a similar trend to figure 7, with most regions falling between the range of -1 to 0. There appears to be a clear trend in the southwest of England, Wales and the north of Scotland, with most improved mobile coverage. Figure 9 shows that most of the areas where coverage is deteriorating are focused around London, which is difficult to see from figure 8.

An issue with this map is that since 0, is included in the -1 to 0 range, it is impossible to tell whether regions in this range have seen a decrease in coverage, or if they have just remained at 0 over the period. Accordingly, conclusions drawn from this band will be unreliable, so the focus is principally on trends from other bands.

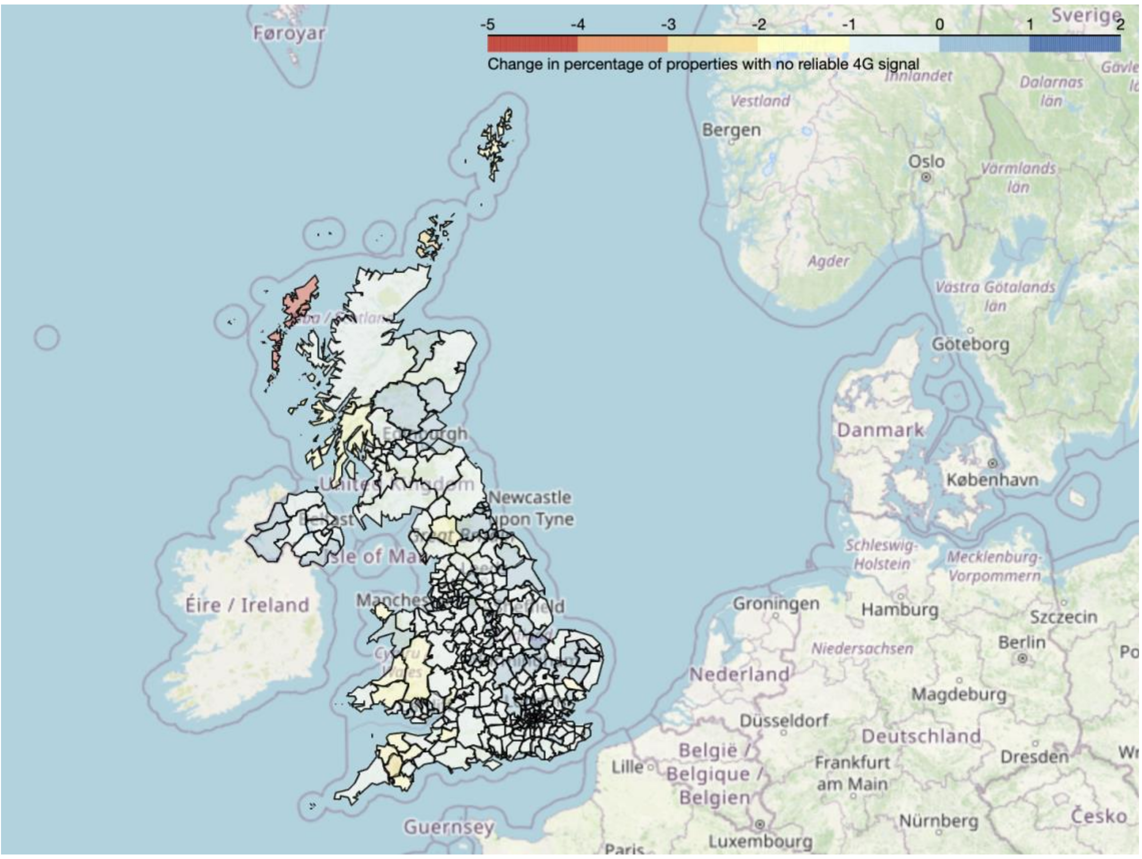

Figure 8: UK map showing change in percentage of premises with no reliable 4G coverage.

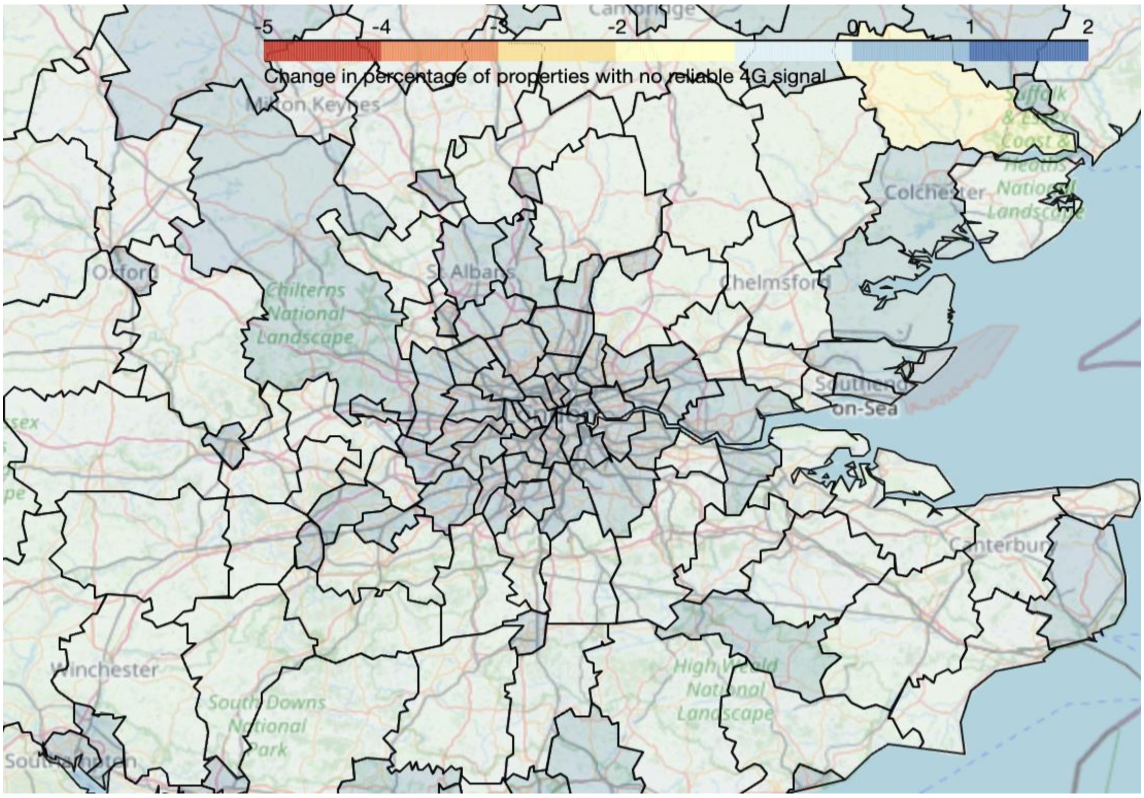

Figure 9: London map showing change in percentage of premises with no reliable 4G coverage.

Overall, whilst UK mobile coverage appears to be improving, there are some clear London regions that are regressing which merit further investigation.

Figure 10 shows the results of the cluster analysis performed in this investigation. There are two clear clusters in this data, one with low values for both variables, and one with higher values. The cluster seems to show a trend of areas that have less than 2% of premises below the USO with no reliable 4G signal, and another cluster showing points with larger percentages.

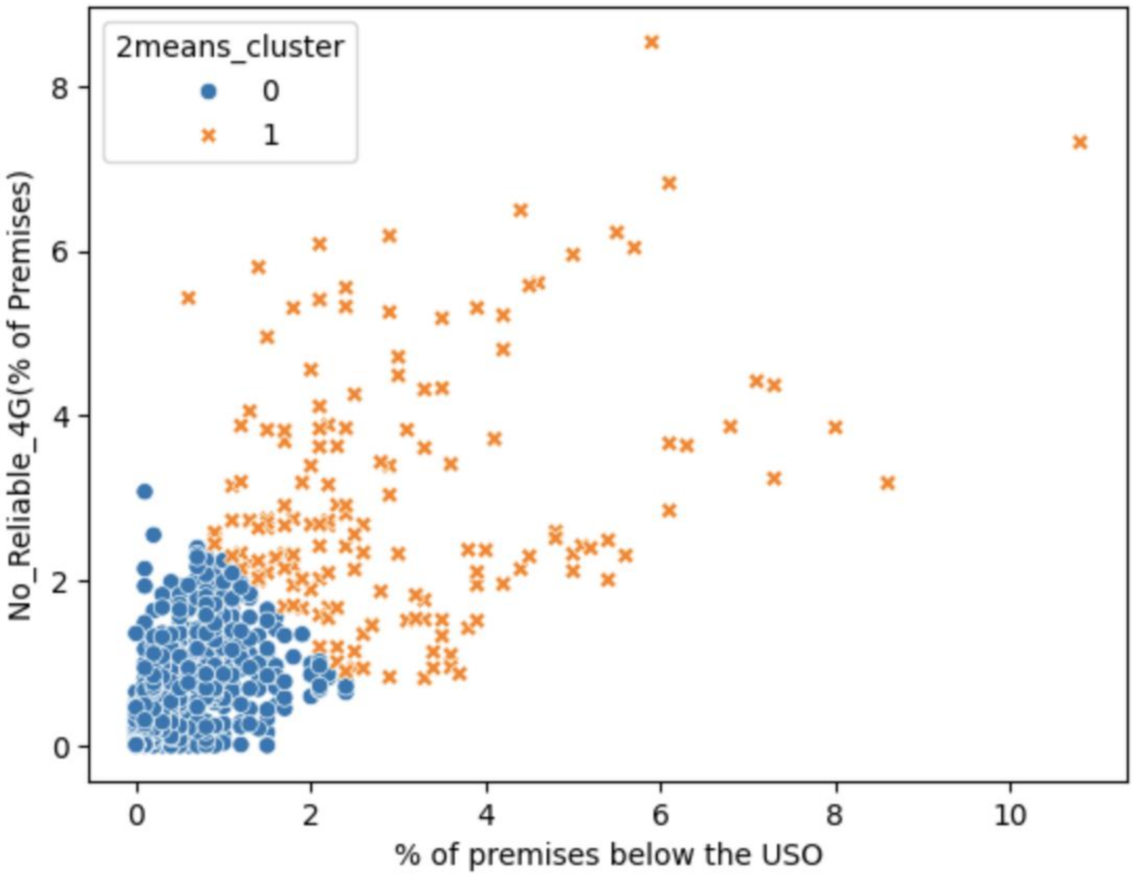

Figure 10: Cluster plot of percentage of premises below the USO vs percentage of premises with no reliable 4G

Table 1 suggests that there has been an increase in the number of premises with poor broadband and mobile coverage, which would be contrary to the investigation earlier suggesting mobile coverage is improving. However, the difference is small and this variance could be explained in slight differences in cluster placement as opposed to a trend of worsening coverage. This is supported further by figures 1 and 4 which show that the cluster coefficient for some of the points in the high-high clusters are much closer to 0.

Dataset	Cluster 0 (Low-Low)	Cluster 1 (High-High)
2019	337	34
2023	325	49
Full	1705	160

Table 1: Counts for the number of points in each cluster based on the dataset used

Pearson's R test gave an R-value of 0.764 with a p-value of 0.0 suggesting a strong positive relationship between these variables, and that the results are statistically significant.

The cluster analysis shows a clear correlation between areas with poor broadband coverage also having poor mobile coverage. This highlights a clear regional inequality in the UK with combined poor performance in both coverages. Whilst the number of areas that fall into this clusters could be varied, there is a clear linear relationship between the two variables, which is further supported by the results from Pearson's R test.

This investigation does show clear trends, however, there are several potential confounding factors. The collection of the mobile data could be impacted by the collection time. If some areas were measured during peak usage hours, they may show lower results than those tested in low use times. This risk is mitigated as the data is sourced by Ofcom and therefore should use a reliable process. Additionally, the phone used in testing could impact results, with newer phones able to access more networks (Jasso, 2023). This could have impacted results as using the same phone for all premises would be almost impossible.

Conclusions

This investigation has shown that between 2019 and 2023 mobile coverage in the UK has been improving, although not universally. The relationship between regions with poor metrics on both coverages was investigated with a clear linear relationship observed. Whilst only around 10% of local authorities in 2023 fell into the cluster for poor coverage, it is vital these inequalities are addressed given the criticality of such coverage.

Whilst this report has highlighted these inequalities, further work needs to be done to identify the specific regions to address the problem. Additionally, this report has not investigated potential underlying reasons behind these inequalities, to address the root cause of the problem, which is an area for future work.

They were limitations to the investigation, noted within the report, including changes to the datasets, and specific problems with the types of analysis used, however, the impact on the strength of conclusions were addressed.

Reflections

The cleaning and storage of the broadband coverage dataset is described in TMA02.

The mobile data was relatively easy to work with, and since there were no problems with data size, it was imported and cleaned using pandas. I chose to remove rows for outdated local authorities from the dataset, as there were few data entries, and the size of the dataset would be sufficient for accurate decisions. I could have chosen to aggregate the rows into the updated councils. However, as the data was stored as percentages, with no counts, aggregation would have been too difficult and introduced greater risks of incorrect data. Whilst my decision may have limited the reliability of my analysis slightly, I thought it was wiser than risking incorrect data that could skew the results.

To deal with the null values, I investigated the relevant situations. All null values appeared where the total for the other columns had reached 100%. This indicated null values where being used to represent 0, so I changed them to this. This had the added benefit of allowing me to use averages without having to worry about math errors. I considered leaving all null values as null, but that would have meant more missing data making results less reliable.

After cleaning the datasets using a similar process to TMA02, I chose to store the data in MongoDB as this allowed the use of aggregation pipelines. This made it easier to create smaller dataframes, separated by year, and with averaged columns. In hindsight, the pipelines often required difficult and time-consuming code. It may have been easier to use panadas throughout to create the smaller dataframes used for producing data representations.

Combining the datasets was not difficult but required using local authority codes as opposed to names because of inconsistencies in naming conventions. I joined on year to have one row for each local authority and year, as opposed to multiples, to facilitate plotting graphs. This decision did make it harder to see which points referred to which areas as I'm unfamiliar with the codes, but this could be solved with googling.

K-means clustering was chosen to allow insight into distribution of the data, to identify trends and to compare the change over time. It didn't make sense to use k-NN predictions as there were no labels for classification.

(2994 words)

References

Building Digital UK (2022) 'Benefits of Rural Mobile Coverage' Available at: https://www.gov.uk/government/publications/benefits-of-rural-mobile-coverage/benefits-of-rural-mobile-coverage#appendix (Accessed 30 May 2024)

Jasso, A. (2023) *Which Cell Phone Gets The Best Reception?* Available at: https://www.wilsonamplifiers.com/blog/which-type-of-smartphone-gets-the-best-reception/#:~:text=Newer%20phones%20do%20tend%20to,the%20best%20and%20fastest%20signals. (Accessed 1 June 2024)

Ofcom (2023a) *Mobile Matters 2023 report*. Available at: https://www.ofcom.org.uk/__data/assets/pdf_file/0018/264600/mobile-matters-2023-report.pdf (Accessed 30 May 2024)

Ofcom (2023b) *Universal service obligations (broadband and telephony).* Available at: https://www.ofcom.org.uk/phones-telecoms-and-internet/information-for-industry/telecoms-competition-regulation/general-conditions-of-entitlement/universal-service-obligation (Accessed 30 May 2024)

Ofcom (2023c) *Connected Nations and infrastructure reports*. Available at: https://www.ofcom.org.uk/research-and-data/multi-sector-research/infrastructure-research (Accessed: 29 May 2024).

The National Archive, *Open Government Licence v3.0*. Available at: https://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/ (Accessed 29 May 2024)

Valentín-Sívico, J., Canfield, C., Low, S., Gollnick, C. (2023) 'Evaluating the impact of broadband access and internet use in a small underserved rural community' *Telecommunications Policy,* vol 47, Issue 4. Available at: https://doi.org/10.1016/j.telpol.2023.102499 (Accessed 30 May 2024)

Appendix 1: Notebooks

The notebooks used for processing and analysing the two datasets can be found below. They should be run in the order presented.

Notebook	Contents
J7526527_q1_TMA02_EMA.ipynb	Preprocessing, cleaning and storage of the broadband coverage dataset.
J7526527_project_diary_EMA.ipynb	Preprocessing, cleaning and storage of the mobile network coverage dataset. Investigation into the mobile network coverage dataset. Combining the mobile network coverage and broadband coverage datasets. Investigation into the combined dataset including k-means clustering algorithm.