A demographic deep dive into internet adoption

Annex: Analysis of Ofcom's Technology Tracker data 2024

Contents

Section

Annex

1. Overview

1.1. Background

This report summarises findings from exploratory analysis of Ofcom's Technology Tracker, July 2024[1]; a nationally representative survey of 4,000 UK adults (aged 16 and above).

The analysis explored demographic characteristics of groups of the UK adult population that are either offline entirely or those that have limited methods of accessing internet connectivity. Logistic regression analysis was used to assess the relative strength of correlation between methods of internet access and demographic characteristics. This helps to show the characteristics which index highly in the following groups:

i) Non-Internet Users;
ii) Mobile Data only Users; and
iii) External Connectivity only Users

2. Considerations and notes on analysis

The Technology Tracker is a general population survey using randomly selected cluster sampling to measure the incidence of technologies used in the household of an individual/ by individuals themselves, in the UK, aged 16+. Details on the mode and sample for the Technology Tracker can be found in the associated technical report. Sampling and weighting are based on nations and the UK at an overall level, and not set for individual regions to be representative.

Given the low numbers of respondents in the groups analysed in this report, we must be particularly mindful that small differences in our respondent base may not reflect real differences in the whole population. Where possible we have conducted statistical significance testing.

In particular, the outputs from our modelling work in Section 4 are based on a low number of responses – and more work should be done in the future to combine years of data together to add to confidence in the findings. Both modelling approaches produced results which were broadly in line with each other.

[1] Data collection occurred from 8 January 2024 to 30 April 2024

2. Methodology

2.1 Core Objective

The objective of this analysis is look at distinct groups of UK adults (aged 16 and above) based on their type of internet access method and usage, using data from Ofcom's Technology Tracker 2024. This analysis examines the size of these groups and explores their demographic characteristics, including age, housing tenure, employment status, socio-economic classification, and the presence of limiting or impacting conditions. By leveraging both descriptive and predictive methodologies, the analysis seeks to understand the characteristics of those with limited internet access use and non-internet use.

2.1 Overview of Methodology

The methodology used for the analysis being described in this report is outlined below.

Data Preparation and Weighting

Data Preparation: blank values are explicitly flagged as "NA", categorical data is converted into consistent codes, and inconsistencies in codes are resolved to enable analysis to be conducted. One respondent was removed from the overall sample as they had not responded to enough questions to be included in the analysis

Weighting: Survey weights are applied to ensure the analysis accurately reflects the UK population distribution (NB this was already included on the Technology Tracker dataset).

Phase 1: Descriptive Analysis

This phase aims to provide a detailed understanding of the percentage of the UK adult population without, or with limited access to in-home internet connectivity, as well as the demographic variations (such as age, housing tenure, employment status, socio-economic group, and limiting or impacting conditions).

Employing descriptive analysis techniques, such as frequency and percentage calculations, establishes a clear baseline of these groups' size and characteristics. Cross-tabulations are specifically used to compare demographic subgroups within each defined group (i.e. Non-Internet Users, Mobile Data only Users, and External Connectivity only Users).

This allows for a detailed examination of how different demographic factors, such as age groups or housing tenure, vary within and across these internet access groups. For example, whether individuals with different housing tenures have distinct levels of internet access. By analysing these subgroup differences, the report provides insights into how socio-demographic characteristics intersect with levels of internet access.

Framework for analysis

The analysis has focussed on three groups, which are mutually exclusive. The following definitions are used:

1. **Non-Internet Users:** to describe those people who do not access the internet. This means those without broadband or mobile-data internet in the home and who don't access the internet outside of the home. **Mobile Data only Users:** to describe those people who don't have broadband at home and only access the internet at home through their own mobile data. **External Connectivity only Users:** to describe those people who don't have broadband at home, do not access the internet through their own mobile data and only access the internet through an external connection (e.g. public wifi, etc).

Annex B describes the definition used for the groups and the questions used to assign respondents to each group.

Size of sample in each group

Table 1 presents the unweighted sample size of each group.

Table 1: Sample sizes per group

Technology Tracker data	Number of respondents in this group	Number of respondents not in this group	Total number of respondents in Technology Tracker
Non-Internet Users	145	3855	4000
Mobile data only Users	130	3870	4000
External connectivity only Users	42	3958	4000

Phase 2: Predictive Modelling

This phase aimed to understand the factors influencing internet access. Techniques used included:

Logistic Regression with Recursive Feature Elimination (RFE) to isolate statistically significant predictors of belonging to an internet access user group (binary outcome: excluded = 1, not excluded = 0). RFE systematically eliminates less significant variables and addresses multicollinearity (where two or more variables are correlated). As this model had a good model fit across all three groups and could produce odds ratios which are more straightforward for users to understand we have used the figures from this model in the final report.

Random Forest was employed to capture complex, non-linear relationships between predictors and the outcome, leveraging feature importance (measured through Gini impurity) and cross-validation for robust analysis. A hybrid sampling approach, combining Random Under-Sampling (RUS) and Synthetic Minority Oversampling Technique (SMOTE), was applied to address class imbalances, with synthetic samples weighted to prevent overfitting. Key steps included data preprocessing for standardisation and missing value handling, as well as multicollinearity checks using Variance

Inflation Factor (VIF). Models, including Logistic Regression and Random Forest, were trained on balanced datasets and evaluated using Precision, Recall, F1-Score, and ROC-AUC metrics. These evaluation metrics provided a comprehensive understanding of model performance, balancing accuracy and interpretability.

3 Descriptive analysis

3.1 Overall estimated size of groups

- Non-Internet Users: Approximately 3.4% (1.9m) of the UK adult population is completely offline both inside and outside the home;
- Mobile Data only Users: Approximately 3.8% of the UK adult population only uses mobile data to access internet connectivity;
- External Connectivity only Users: 0.9% rely on 'external connectivity'

3.2 Age

3.2.1 Distribution by age

Table 2: Internet access method by Age Groups, highlighting statistically significant differences

Group	Non-Internet Users	Mobile Data only	External connectivity	All Respondents
16–17 years	0 (0.0%)	10 (6.8%)	2 (5.7%)	109 (2.7%)
18–24 years	15 (11.1%)	30 (19.7%)	14 (40.0%) ***	418 (10.4%)
25–34 years	12 (8.9%)	27 (17.8%)	4 (11.4%)	650 (16.2%)
35–44 years	11 (8.1%)	18 (11.8%)	7 (20.0%)	727 (18.2%)
45–54 years	3 (2.2%)	32 (21.1%)	3 (8.6%)	545 (13.6%)
55–64 years	18 (13.3%)	14 (9.2%) ***	1 (2.9%) ***	677 (16.9%)
65–74 years	18 (13.3%) ***	17 (11.2%)	2 (5.7%)	504 (12.6%)
75–84 years	30 (22.2%) ***	4 (2.6%)	2 (5.7%)	290 (7.2%)
85+ years	28 (20.7%) ***	0 (0.0%)	0 (0.0%)	81 (2.0%)
TOTAL	136	151	34	4000

a) P-values are derived from chi-square tests assessing differences in internet access among age groups.
*b) Significance levels are denoted as follows: p < 0.05, denoted by *, p < 0.01, denoted by **, p < 0.001, denoted by ****
c) Pairwise comparisons are corrected for multiple testing using Bonferroni adjustment where applicable.
d) Due to rounding and the nature of weighting, the sum of weighted values may differ slightly from the unweighted totals.

3.3 Socio-Economic Group (SEG)

Table 3: Internet access method by SEG, highlighting statistically significant differences

Group	Non-Internet Users	Mobile Data only	External connectivity	All Respondents
A	4 (2.9%)	5 (3.3%)	0 (0.0%)	190 (4.8%)
B	6 (4.4%)	16 (10.5%) **	4 (11.8%)	713 (18.1%)
C1	27 (19.9%)	27 (17.8%)	10 (29.4%)	1287 (32.7%)
C2	25 (18.4%)	51 (33.6%) ***	7 (20.6%)	850 (21.6%)
D	20 (14.7%)	18 (11.8%) **	5 (14.7%)	313 (7.9%)
E	52 (38.2%) ***	31 (20.4%)	8 (23.5%)	588 (14.9%)
TOTAL	136	151	34	4000

a) *P-values are derived from chi-square tests assessing differences in internet access among SEG groups.*
b) *Significance levels are denoted as follows: $p < 0.05$, denoted by *, $p < 0.01$, denoted by **, $p < 0.001$, denoted by ****
c) *Pairwise comparisons are corrected for multiple testing using Bonferroni adjustment where applicable. Due to rounding and the nature of weighting, the sum of weighted values may differ slightly from the unweighted totals.*

3.4 Housing Tenure

Table 4: Internet access method by Housing tenure groups, highlighting statistically significant differences

Group	Non-Internet Users	Mobile Data only	External connectivity	All respondents
Being bought on mortgage	5 (3.6%)	29 (19.2%)	5 (14.7%) **	1133 (28.5%)
Owned outright by household	44 (32.1%) ***	20 (13.2%)	7 (20.6%)	1226 (30.9%)

Group	Non-Internet Users	Mobile Data only	External connectivity	All respondents
Rented from Local Authority/ Housing Association/ Trust	51 (37.2%) ***	46 (30.5%) ***	12 (35.3%)	723 (18.2%)
Rented from private landlord	10 (7.3%)	41 (27.2%) ***	7 (20.6%)	596 (15.0%)
Don't know	13 (9.5%) ***	3 (2.0%)	0 (0.0%)	58 (1.5%)
Prefer not to say	14 (10.2%) **	8 (5.3%)	3 (8.8%)	235 (5.9%)
TOTAL	136	151	34	4000

a) *P-values are derived from chi-square tests assessing differences in internet access among Housing tenure groups.*
b) *Significance levels are denoted as follows: p < 0.05, denoted by *, p < 0.01, denoted by **, p < 0.001, denoted by ****
c) *Pairwise comparisons are corrected for multiple testing using Bonferroni adjustment where applicable.*
d) *Other (type) not included in the percentages.*
Due to rounding and the nature of weighting, the sum of weighted values may differ slightly from the unweighted totals.

3.5 Employment Status

Table 5: Internet access method by Employment Status groups, highlighting statistically significant differences

Group	Non-Internet Users	Mobile Data only	External Connectivity	All respondents
Student	0 (0.0%)	16 (10.6%)	4 (11.8%)	221 (5.5%)
Full-time responsibility for home/family	8 (5.9%) *	4 (2.6%)	3 (8.8%)	157 (3.9%)
In full-time employment	13 (9.6%) ***	57 (37.1%)	16 (47.1%)	1869 (46.7%)

Group	Non-Internet Users	Mobile Data only	External Connectivity	All respondents
In part-time employment	8 (5.1%)	21 (13.9%)	5 (14.7%)	486 (12.2%)
Retired	84 (61.8%) ***	21 (13.9%)	3 (8.8%)	946 (23.7%)
Unemployed	15 (11.0%) ***	31 (20.5%) ***	3 (8.8%)	253 (6.3%)
Prefer not to say	9 (6.6%)	2 (1.3%)	0 (0.0%)	68 (1.7%)
TOTAL	136	151	34	4000

a) P-values are derived from chi-square tests assessing differences in internet access among Employment status groups.

*b) Significance levels are denoted as follows: $p < 0.05$, denoted by *, $p < 0.01$, denoted by **, $p < 0.001$, denoted by ****

c) Pairwise comparisons are corrected for multiple testing using Bonferroni adjustment where applicable. Due to rounding and the nature of weighting, the sum of weighted values may differ slightly from the unweighted totals.

3.6 Benefit Type

Table 6: Internet access method by type of welfare benefit, highlighting statistically significant differences

Benefit Type	Non-Internet Users	Mobile Data only	External connectivity	All respondents
Disability Living Allowance	1 (0.4%)	0 (0.0%)	0 (0.0%)	139 (3.3%)
Employment and Support Allowance (ESA)	1 (0.4%)	0 (0.0%)	0 (0.0%)	69 (1.6%)
Income Support	27 (20.0%) **	27 (18.0%) **	4 (11.8%)	33 (0.8%)
Income-based Jobseeker's Allowance	1 (0.4%)	0 (0.0%)	1 (2.9%)	20 (0.5%)
Pensions Credit	18 (13.5%) **	1 (0.7%)	0 (0.0%)	141 (3.4%)

Benefit Type	Non-Internet Users	Mobile Data only	External connectivity	All respondents
Personal Independence Payment (PIP)	2 (1.5%)	3 (1.9%)	3 (8.8%)	184 (4.4%)
Universal Credit (with other earnings)	4 (3.0%)	11 (7.0%)	2 (5.8%)	168 (4.0%)
Universal Credit (without other earnings)	2 (1.5%)	16 (10.6%) **	1 (2.9%)	130 (3.1%)
Don't Know	0 (0.0%)	0 (0.0%)	1 (2.9%)	2758 (65.7%)
None – do not receive any benefits	63 (46.2%) **	82 (54.5%) **	22 (64.7%) **	158 (3.8%)
Prefer not to say	8 (6.1%)	8 (5.5%)	0 (0.0%)	359 (8.6%)
Something else	10 (7.1%) **	3 (1.9%)	0 (0.0%)	37 (0.9%)
TOTAL	136	151	34	4000

a) *P-values are derived from chi-square tests assessing differences in internet access among Benefits groups.*
b) *Significance levels are denoted as follows: p < 0.05, denoted by *, p < 0.01, denoted by **, p < 0.001, denoted by ****
c) *Pairwise comparisons are corrected for multiple testing using Bonferroni adjustment where applicable.*
d) *Due to rounding and the nature of weighting, the sum of weighted values may differ slightly from the unweighted totals.*

3.7 Region

Table 7: Internet access method by Region, highlighting statistically significant differences

Region	Non-Internet Users	Mobile Data only	External connectivity	All respondents
East Midlands	10 (6.9%)	10 (6.4%)	0 (0.0%)	292 (7.3%)
East England	11 (8.0%)	3 (2.0%) *	2 (5.9%)	376 (9.4%)
London	20 (14.9%)	29 (19.5%) **	1 (2.5%)	520 (13.0%)

Region	Non-Internet Users	Mobile Data only	External connectivity	All respondents
North East England	1 (0.9%)	22 (14.5%) **	2 (5.9%)	160 (4.0%)
North West England	8 (6.2%)	20 (13.0%)	2 (5.9%)	444 (11.1%)
Northern Ireland	4 (2.8%)	2 (1.4%) **	3 (8.2%)	108 (2.7%)
Scotland	12 (9.1%)	19 (12.4%)	4 (11.4%)	332 (8.3%)
South East England	12 (8.8%) *	5 (3.2%) **	9 (26.9%) *	552 (13.8%)
South West England	9 (6.2%)	3 (2.0%) **	4 (11.4%)	348 (8.7%)
Wales	11 (8.0%)	5 (3.2%) **	1 (1.5%)	188 (4.7%)
West Midlands	17 (12.4%)	17 (11.1%)	6 (16.5%) *	352 (8.8%)
Yorkshire and The Humber	22 (15.9%) **	17 (11.1%)	1 (3.3%)	328 (8.2%)
TOTAL	136	151	34	4000

a) *P-values are derived from chi-square tests assessing differences in internet access among Region.*
b) *Significance levels are denoted as follows: $p < 0.05$, denoted by *, $p < 0.01$, denoted by **, $p < 0.001$, denoted by ****
c) *Pairwise comparisons are corrected for multiple testing using Bonferroni adjustment where applicable.*
d) *Due to rounding and the nature of weighting, the sum of weighted values may differ slightly from the unweighted totals.*

3.8 Urbanity

Table 8: Internet access method by Urbanity, highlighting statistically significant differences

Urban/Rural	Non-Internet Users	Mobile Data only	External connectivity	All respondents
Rural	19 (14.2%)	25 (16.8%)	13 (36.9%)	722 (18.1%)
Urban	117 (85.8%)	126 (83.2%)	22 (63.1%)	3278 (81.9%)
TOTAL	136	151	34	4000

a) *P-values are derived from chi-square tests assessing differences in internet access among Urbanity.*
b) *Significance levels are denoted as follows: p < 0.05, denoted by *, p < 0.01, denoted by **, p < 0.001, denoted by ****
c) *Pairwise comparisons are corrected for multiple testing using Bonferroni adjustment where applicable.*
d) *Due to rounding and the nature of weighting, the sum of weighted values may differ slightly from the unweighted totals.*

3.9 Locale

Table 9: Internet access method by Locale, highlighting statistically significant differences

Locale	Non-Internet Users	Mobile Data only Users	External connectivity	All respondents
Large City	22 (15.9%)	38 (25.0%) *	1 (2.8%)	483 (12.1%)
Smaller City or Large Town	22 (16.1%)	24 (15.6%)	8 (24.8%)	759 (19.0%)
Medium Town	43 (31.1%)	43 (28.5%) *	9 (25.9%)	1232 (30.8%)
Small Town within 10 miles from a settlement with 15K+ pop	20 (14.8%)	18 (11.7%)	2 (6.8%)	707 (17.7%)
Small Town more than 10 miles from a settlement with 15K+ pop	11 (7.8%)	4 (2.3%)	1 (2.8%)	126 (3.2%)
Rural Area within 10 miles from a settlement with 15K+ pop	14 (10.2%)	13 (8.9%)	10 (28.0%)	529 (13.2%)
Rural Area more than 10 miles from a settlement with 15K+ pop	5 (4.0%)	12 (7.9%)	3 (8.8%)	164 (4.1%)
TOTAL	136	151	34	4000

a) *P-values are derived from chi-square tests assessing differences in intent access among Locale.*

b) *Significance levels are denoted as follows: p < 0.05, denoted by *, p < 0.01, denoted by **, p < 0.001, denoted by ****
c) *Pairwise comparisons are corrected for multiple testing using Bonferroni adjustment where applicable.*
d) *Due to rounding and the nature of weighting, the sum of weighted values may differ slightly from the unweighted totals.*

3.10 People in the household

Table 10: Internet access method by the Number of People in the Household, highlighting statistically significant differences

Number of people in household	Non-Internet Users	Mobile data only	External connectivity	All respondents
1	73 (53.3%) *	38 (25.2%)	7 (20.6%)	769 (19.2%)
2	37 (27.0%) *	47 (31.1%)	19 (55.9%)	1456 (36.4%)
3	9 (6.6%) *	29 (19.2%)	4 (11.8%)	736 (18.4%)
4	16 (11.7%) *	21 (13.9%)	2 (5.9%)	668 (16.7%)
5+	2 (1.5%) *	16 (10.6%)	2 (5.9%)	371 (9.3%)
TOTAL	136	151	34	4000

a) *P-values are derived from chi-square tests assessing differences in internet access among household People.*
b) *Significance levels are denoted as follows: p < 0.05, denoted by *, p < 0.01, denoted by **, p < 0.001, denoted by ****
c) *Pairwise comparisons are corrected for multiple testing using Bonferroni adjustment where applicable.*
d) *Due to rounding and the nature of weighting, the sum of weighted values may differ slightly from the unweighted totals.*

4. Predictive Modelling analysis

4.1 Model Development and Evaluation

This analysis utilised Logistic Regression and Random Forest models to identify key factors influencing in-home internet connectivity. The models are trained and validated using an imbalanced dataset, with additional steps to address class imbalance through hybrid sampling techniques (Random Under-sampling and SMOTE).

4.2 Non-Internet Users

Logistic Regression with Recursive Feature Elimination (RFE)

The logistic regression model includes the selection of 35 features through RFE. Key findings include the values with $p < 0.05$:

Final Model Equation:

y=−2.9482 + (2.0857 × Age - 85 years or over) + (1.7410 × Impacting or limiting condition - Learning or cognitive abilities) + (−0.8384 × People in Household - 1) + (1.1792 × Employment Status - Retired) + (1.5037 × Benefits - Don't know) + (0.7586 × Housing Tenure -Rented from Local Authority / Housing Association/ Trust) + (1.4022 × Employment Status – Full-time responsibility for the home / family) + (0.9304 × Region - Yorkshire and The Humber) + (0.7060 × Age – 75-84 years) + (−0.7308 × Financial Status - Doing well) + (−1.1348 × Impacting or limiting condition - Breathing) + (1.2750 × Benefits - Income Support)

Key Predictors:

Age:
- 85 years or over: Households with individuals aged 85+ are 8.05 times more likely to be Non-Internet Users when controlling for other factors (Odds Ratio = 8.05, Coefficient = 2.0857).
- 75–84 years: Similar increased risk with an Odds Ratio of 2.03 (Coefficient = 0.7060).

Impacting or limiting conditions:
- Respondents who had a learning or cognitive abilities impacting or limiting condition: Significantly increases the chance of being a Non-Internet User (Odds Ratio = 5.70, Coefficient = 1.7410).
- Breathing impacting or limiting condition: Decreases the chance (Odds Ratio = 0.32, Coefficient = -1.1348).

Household Composition:
- Single-person households: Significantly reduces chance of being a Non-Internet User (Odds Ratio = 0.43, Coefficient = -0.8384).

Employment Status:
- Retired individuals: Are more likely to be Non-Internet Users, with an Odds Ratio of 3.25 (Coefficient = 1.1792).

- Full-time family/home responsibilities: Also, more likely to be Non-Internet Users, with an Odds Ratio of 4.06 (Coefficient = 1.4022).

Region:
- Yorkshire and The Humber: Higher likelihood of non-internet use (Odds Ratio = 2.54, Coefficient = 0.9304).

Housing Tenure:
- Renting from Local Authority/Housing Associations: Increases the likelihood of non-internet use (Odds Ratio = 2.14, Coefficient = 0.7586).

Financial Status:
- "Doing well" financially: Reduces the likelihood of non-internet use by approximately 50% (Odds Ratio = 0.48, Coefficient = -0.7308).

Benefits:
- Income Support: Increases the likelihood of non-internet use (Odds Ratio = 3.58, Coefficient = 1.2750).
- Don't know: Households uncertain about their benefits status had an increased likelihood to be Non-Internet Users (Odds Ratio = 4.50, Coefficient = 1.5037).

Random Forest Model

The Random Forest model provides an alternative method to evaluate the relative importance of demographic features.

Key Predictors:

- **Employment Status - Retired**: With the highest importance score of **0.164**, being retired is the most significant predictor of non-internet use in this model.
- **People in the household (1)**: Living in a single-person household is the next strongest predictor, with an importance score of **0.137**.
- **Socio-Economic Group (SEG E)**: Individuals in socio-economic group E are more likely to be Non-Internet Users, with an importance score of **0.107**.
- **Housing Tenure**:
  - **Rented from Local Authority/Housing Association/Trust**: This tenure significantly contributes to non-internet use, with an importance score of **0.051**.
  - **Being bought on mortgage**: Also relevant, with an importance score of **0.038**.
- **Age Groups**:
  - **75–84 years**: Older individuals in this age group are influential predictors, with an importance score of **0.055**.
  - **85+ years**: This group also contributes to non-internet use, with an importance score of **0.034**.
- **Socio-Economic Group (SEG C1)**: Individuals in this group have a moderate likelihood of non-internet use, reflected in an importance score of **0.029**.
- **Region**:
  - **Wales**: Residing in Wales significantly increases the likelihood of non-internet use, with an importance score of **0.029**.
  - **Yorkshire and The Humber**: Living in this region is associated with a higher likelihood of non-internet use, with an importance score of **0.026**.
- **Financial Situation**:
  - **Struggling**: Financial difficulties contribute to non-internet use, with an importance score of **0.029**.

    - **Doing well**: Managing finances well has a notable impact, with an importance score of **0.030**.
- **Impacting or limiting condition**:
    - **Learning or cognitive abilities**: This impacting or limiting condition is a significant factor, with an importance score of **0.024**.
    - **Hearing**: Hearing impacting or limiting conditions also contribute, with an importance score of **0.017**.
- **Having children (Yes/No)**: The presence of children affects the likelihood of non-internet use, with importance scores of **0.036** (Yes) and **0.033** (No).

Model Fit and Performance

The performance of three models—Logistic Regression (before and after hybrid sampling) and Random Forest (after hybrid sampling)—was evaluated to assess their effectiveness in predicting non-internet use. The results are summarised below:

Logistic Regression (Before Oversampling)

- Pseudo R^2: The model explains 35.2% of the variance in non-internet use, indicating a moderate-low fit.
- Precision: Achieved a score of 28%, meaning that when predicting non-internet use, it is correct in 28% of cases.
- Recall: 46% of Non-Internet Users are identified, showing moderate sensitivity to the excluded class.
- AUC Score: A score of 0.79 suggests good discrimination between included and excluded respondents.

Logistic Regression (After Hybrid Sampling)

- Precision: Drops to 18%, as the model becomes less precise in its predictions after addressing class imbalance.
- Recall: Significantly improves to 68%, indicating a higher ability to identify Non-Internet Users.
- AUC Score: A slight increase to 0.83, reflects an improvement in overall performance in distinguishing between the two classes.

Random Forest (After Hybrid Sampling)

- Precision: Matches the precision of the Logistic Regression before oversampling at 28%.
- Recall: Lower than the oversampled Logistic Regression, identifying 32% of Non-Internet Users.
- AUC Score: Achieves the highest score of 0.84, indicating the best ability among the models to distinguish Non-Internet Users.

Table 12: Model metrics across Logistic Regression and Random Forest before and after sampling for Non-Internet Users

Metric	Logistic Regression (Before Oversampling)	Logistic Regression (After Hybrid Sampling)	Random Forest (After Hybrid Sampling)
Pseudo R^2	0.352		
Precision	28%	18%	28%
Recall	46%	68%	32%
AUC Score	0.79	0.83	0.84

4.3 Mobile Data only Users

Logistic Regression with Recursive Feature Elimination (RFE)

The logistic regression model selects 35 features through RFE. Key findings include the values with $p < 0.05$:

Final Model Equation:

Y = −2.4758 + (2.0094× Region-North East) + (1.0886 × Housing Tenure – Rented from private landlord) + (0.8504 × Housing Tenure – Rented from Local Authority/ Housing Association/ Trust) + (-0.7059 * People in household - 1) + (1.0737 x Locale - Rural Area more than 10 miles from a settlement with 15K+ pop) + (0.7851 × Benefits – Universal Credit (without other earnings)) + (0.6787 × Benefits – Prefer not to say

Key Predictors:

Region:

- **North East England:** Adults (over 16+) in North East England are about 7.46 times more likely to be mobile data only users (Odds Ratio = 7.46, Coefficient = 2.0094).

Household Composition:

- **Single-person households** are 51% less likely to be mobile data only users (Odds Ratio = 0.49, Coefficient = -0.7059).

Housing Tenure:

- **Rented from private landlord:** Those people renting privately are nearly 3 times more likely to be mobile data only users (Odds Ratio = 2.97, Coefficient = 1.0886).
- **Rented from Local Authority/ Housing Association/ Trust:** This type of renter is about 2.34 times more likely to be a mobile data only user (Odds Ratio = 2.34, Coefficient = 0.8504).

Locale (**Rural Area more than 10 miles from a settlement with 15K+ pop**):

People living in rural areas more than 10 miles from a settlement (with a population of over 15,000) are nearly 3 times more likely to be mobile data only users (Odds Ratio = 2.93, Coefficient = 1.0737).

Benefits:

- **Universal Credit (without other earnings):** Increases the likelihood of mobile data only use by 2.19 times (Odds Ratio = 2.19, Coefficient = 0.7851).
- **Prefer not to say:** Adults (16+) who are unwilling to disclose their benefits status are about 2 times more likely to be mobile data only users (Odds Ratio = 1.97, Coefficient = 0.6787).

Random Forest Model

The Random Forest model provides an alternative approach to determine feature importance.

Key Predictors:

- **Socio-Economic Group (B):** With the highest importance score of 0.164, being retired is the most significant predictor of mobile data only use in this model.
- **Region**:
  - **North East England**: Living in the North East is a critical predictor, with an importance score of 0.065.
  - **South East England**: Residing in the South East contributes moderately to mobile data only (Importance = 0.045)
  - **Northern Ireland:** Importance score of 0.042
  - **London:** Importance score of 0.034.
- **Housing Tenure**:
  - Rented from Local Authority/ Housing Association/ Trust: Importance score of 0.063.
  - Rented from private landlord: Importance score of 0.060.
- **Employment Status - Retired:** Retirement status strongly impacts mobile data only use, with an importance score of 0.057.
- **People in the household (3 or more**): Larger households significantly influence the likelihood of mobile data only use **(Importance = 0.048)**.
- **Age Groups:**
  - **55 – 64 years**: This age group affects mobile data only use (Importance = 0.046).
  - **18–24 years**: Importance score of 0.028.
- **Financial Situation** (**Struggling)**: Households struggling financially have an importance score of 0.044.
- **Having children (Yes)**: The presence of children affects the likelihood of mobile data only use, with importance scores of 0.043.
- **Locale (Rural Area >10 miles from 15K+ settlement)**: Importance score of 0.039**.**

Model Fit and Performance

The performance of three models—Logistic Regression (before and after hybrid sampling) and Random Forest (after hybrid sampling)—was evaluated to assess their effectiveness in predicting mobile data only use. The results are summarised below:

Logistic Regression (Before Oversampling)

- Pseudo R^2: The model explains 15.6% of the variance in mobile data only use, indicating a low fit.
- Precision: Achieved a score of 9%, meaning that when predicting mobile data only use, it is correct in 9% of cases.
- Recall: 6% of mobile data only user households are identified, showing low sensitivity to the excluded class.

- AUC Score: A score of 0.70 suggests moderate discrimination between included and excluded households.

Logistic Regression (After Hybrid Sampling)

- Precision: Dropped to 6%, reflecting a decrease in the accuracy of predictions mobile data only use after addressing class imbalance.
- Recall: Improved significantly to 46%, indicating a much higher ability to identify mobile data only user households compared to before oversampling.
- AUC Score: Decreased slightly to 0.67, reflecting a moderate ability to distinguish between the two classes.

Random Forest (After Hybrid Sampling)

- Precision: Dropped to 6%, reflecting a decrease in the accuracy of predictions for mobile data only use after addressing class imbalance.
- Recall: Lower than the oversampled Logistic Regression, identifying 23% of mobile data only user households.
- AUC Score: Achieved the same score of 0.67 as the logistic regression with oversample.

Table 13: Model metrics across Logistic Regression and Random Forest before and after sampling for Mobile Data only Users

Metric	Logistic Regression (Before Oversampling)	Logistic Regression (After Hybrid Sampling)	Random Forest (After Hybrid Sampling)
Pseudo R^2	0.156		
Precision	9%	6%	11%
Recall	6%	46%	23%
AUC Score	0.70	0.67	0.67

4.4 External Connectivity only Users

Logistic Regression with Recursive Feature Elimination (RFE)

The logistic regression model selected 35 features through RFE. Key findings include the values with $p < 0.1$ to include more features due to the limitation of features with values less than 0.05:

Final Model Equation:

Y = −5.2200 + (0.9409 × Region - Northern Ireland) + (1.0034 × Housing Tenure - Rented from Local Authority/ Housing Association/ Trust) + (0.7974 × Age - 18 – 24 years) + (−1.2499 × Locale - Small Town within 10 miles from a settlement with 15K+ pop)

Key Predictors:

Region:

- Northern Ireland: Households in Northern Ireland are about 2.6 times more likely to rely on external connectivity to access the internet (Odds Ratio = 2.56, Coefficient = 0.9409).

Housing Tenure:

- **Rented from Local Authority/ Housing Association/ Trust:** Households renting from these entities are about 2.7 times more likely to rely on external connectivity (Odds Ratio = 2.73, Coefficient = 1.0034).

Age Group (18–24 years):

- 18-24 years: Individuals aged 18–24 years are about 2.2 times more likely rely on external connectivity (Odds Ratio = 2.22, Coefficient = 0.7974).

Locale:

- **Small Town within 10 miles from a settlement with 15K+ population:** Households in these areas are 71% less likely to rely on external connectivity (Odds Ratio = 0.29, Coefficient = -1.2499)

Random Forest Model

The Random Forest model provides an alternative perspective on feature importance.

Key Predictors:

- **Age Groups:**
  - 18–24 years: Holds the highest importance score of 0.087, making it the strongest predictor.
  - 55–64 years: Moderately important, with a score of 0.042.
- **Region**:
  - Northern Ireland: A critical predictor with an importance score of 0.077.
  - West Midlands: Significant, with a score of 0.067.
  - South East and South West England: Lower but notable, with scores of 0.030 and 0.021.
- **Housing Tenure**:
  - Rented from Local Authority/ Housing Association/ Trust: Importance score of 0.036.
  - Rented from private landlord: Lower importance, with a score of 0.021.
- **Locale**:
  - Medium-sized towns: Moderate effect with an importance score of 0.026.
  - Smaller cities or rural areas: Lower but notable, around 0.024.
- **Socio-Economic Groups (D and C1/C2):**
  - Group D: Highest influence among socio-economic categories (Importance = 0.039).
  - Groups C1 and C2: Mid-range importance, with scores of 0.028 and 0.025.
- **Employment Status:**
  - Retired: Moderate predictor (Importance = 0.040).
  - Part-time employment: Slightly lower impact (Importance = 0.037).
  - Being a student: Moderately influential (Importance = 0.026).
- **Financial Situation** (**Doing well**): Modest effect (Importance = 0.034).

Model Fit and Performance

The performance of three models—Logistic Regression (before and after hybrid sampling) and Random Forest (after hybrid sampling)—was evaluated to assess their effectiveness in predicting reliance on external connectivity. The results are summarised below:

Logistic Regression (Before Oversampling)

- Pseudo R^2: The model explains 17% of the variance in External Connectivity only, indicating a low fit.
- Precision: Achieves a score of 1%, meaning that when predicting reliance on external connectivity, is correct in 1% of cases.
- Recall: Identified 21% of External Connectivity only are identified, showing low sensitivity to the excluded class.
- AUC Score: A score of 0.49 suggests poor discrimination between included and excluded households.

Logistic Regression (After Hybrid Sampling)

- Precision: Remains at 1%, showing no improvement in accuracy.
- Recall: Improves to 29%, indicating the model now identifies nearly 30% of actual external connectivity reliance.
- AUC Score: Increases to 0.50, reflecting only slight improvement in discrimination.

Random Forest (After Hybrid Sampling)

- Precision: Improves to 40%, meaning the model is correct in 40% of its predictions for external connectivity reliance.
- Recall: Lower than Logistic Regression, identifying 14% of individuals relying on external connectivity.
- AUC Score: Increases to 0.73, showing better discrimination ability than Logistic Regression.

Table 14: Model metrics across Logistic Regression and Random Forest before and after sampling for External Connectivity only

Metric	Logistic Regression (Before Oversampling)	Logistic Regression (After Hybrid Sampling)	Random Forest (After Hybrid Sampling)
Pseudo R^2	0.170		
Precision	1%	2%	40%
Recall	21%	36%	14%
AUC Score	0.49	0.53	0.73

Appendix

Appendix A: Description of Ofcom’s Technology Tracker 2024

The analysis is based on raw data from Ofcom’s Technology Tracker, using the 2024 data from a long-running annual survey designed to monitor UK consumers' attitudes and behaviours toward residential telecommunications, broadcasting, and internet services and technologies. The questionnaire can be found here: Technology Tracker 2024 Questionnaire. A full description of the survey methodology and sample design can be found here: Technology Tracker 2024 Technical Report.

Appendix B: Questions used to filter and identify each group in the Technology Tracker 2024

Group (All groups are exclusive, with no overlaps)	Criteria	Definitions from the Technology Tracker used to identify the group
Non-Internet Users	1. No internet and use at home 2. No smartphone 3. No access to broadband, mobile network, or mobile broadband NB: if there had been overlap, those using the internet outside the home would have been excluded from this group	1. Do not select the option “QE1_Yes – have access and use at home” (indicating no access and use at home). 2. Do not select the option QM2_Yes, and I personally use one (indicating no smartphone ownership and use). 3. Do not select any values for QE7_1, QE7_2, and QE7_3 (indicating no access). If all these conditions are true, the respondent is considered to be ‘Non-Internet User’.
Mobile Data only Internet users People who have no broadband at home but have their own internet access via mobile-data connectivity	1. Has smartphone and uses it 2. Specific internet access 3. No broadband or mobile broadband usage at home	1.Select the option QM2_Yes, and I personally use one (indicating they personally own and use a smartphone). 2.Select the value QE7_2 (indicating they have internet access via smartphone). 3.Do not select any values for QE7_1, QE7_3 and QE7_4. If all these conditions are true, and it does not belong to the ‘Non-internet User’, the respondent is considered to be ‘Mobile data only Internet User’.

External Connectivity only Users People who do not have their own internet access at home but use the internet outside the home (e.g. public WiFi in libraries, internet at work or access someone else's mobile internet)	1. No internet and use at home 2. Internet usage outside home 3. No broadband, mobile broadband, or mobile phone network	1. Do not select the option "QE1_Yes – have access and use at home" (indicating no access and use at home). 2. Any of the columns QE4_1 to QE4_12 are selected (indicating they use outside the home). 3. Do not select any values for QE7_1, QE7_2, and QE7_3 (indicating no access). If these conditions are true, and it does not belong to the 'Non-internet User', the respondent is considered a Limited-Internet User (External connectivity)

Annex C: Logistic Regression Results for Internet Access Types

Internet Access Type	Variables	Significant Coefficients	P-values	Logistic Regression Equation
Non-Internet Users	Age - 85 years or over	2.086	0	y=−2.9482 + (2.0857 × Age - 85 years or over) + (1.7410 × Impacting or limiting condition - Learning or cognitive abilities) + (−0.8384 × People in Household - 1) + (1.1792 × Employment Status - Retired) + (1.5037 × Benefits - Don't know) + (0.7586 × Housing Tenure - Rented from Local Authority/ Housing Association/ Trust) + (1.4022 × Employment Status – Full-time responsibility for the home/ family) + (0.9304 × Region - Yorkshire and The Humber) + (0.7060 × Age – 75-84 years) + (−0.7308 × Financial Status - Doing well) + (−1.1348 × Impacting or limiting condition - Breathing) + (1.2750 × Benefits - Income Support)
	Impacting or limiting condition - Learning or cognitive abilities	1.741	0	
	People in household (1)	-0.838	0.0003	
	Employment Status – Retired	1.179	0.001	
	Benefits – Don't know	1.504	0.002	
	Housing tenure – Rented (Local Authority)	0.759	0.002	
	Employment – Full-time home/family	1.402	0.003	

Internet Access Type	Variables	Significant Coefficients	P-values	Logistic Regression Equation
	Region – Yorkshire & Humber	0.930	0.007	
	Age – 75 – 84 years	0.706	0.015	
	Financial Status – Doing Well	-0.731	0.018	
	Impacting or limiting condition - Breathing	-1.135	0.022	
	Benefits – Income Support	1.275	0.034	
Mobile data- only users	Region – North East	2.009	0	Y = −2.4758 + (2.0094× Region-North East) + (1.0886 × Housing Tenure – Rented from private landlord) + (0.8504 × Housing Tenure – Rented from Local Authority/ Housing Association/ Trust) + (-0.7059 * People in household - 1) + (1.0737 x Locale - Rural Area more than 10 miles from a settlement with 15K+ pop) + (0.7851× Benefits – Universal Credit (without other earnings)) + (0.6787 × Benefits – Prefer not to say
	Housing tenure – Rented (private)	1.089	0	
	Housing tenure – Rented (Local Authority)	0.850	0.001	
	People in household (1)	(1): -0.706	0.004	

Internet Access Type	Variables	Significant Coefficients	P-values	Logistic Regression Equation
	Locale – Rural area	1.074	0.005	
	Benefits - Universal Credit (no earnings)	0.785	0.023	
	Benefits – Prefer not to say	0.679	0.024	
	Age – 18-24 years	0.477	0.06	
	Benefits - Universal Credit (with earnings)	0.691	0.061	
	Region – London	0.612	0.064	
	Financial Status - Struggling	0.429	0.094	
	SEG - B	-0.566	0.095	
	Age – 55 – 64 years	-0.518	0.1	
External Connectivity only Users	Housing tenure – Rented (Local Authority)	1.003	0.038	Y = −5.2200 + (0.9409 × Region - Northern Ireland) + (1.0034 × Housing Tenure - Rented from Local Authority/ Housing Association/ Trust) + (0.7974 × Age - 18 – 24 years) + (−1.2499 × Locale - Small Town within 10 miles from a settlement with 15K+ pop)

Internet Access Type	Variables	Significant Coefficients	P-values	Logistic Regression Equation
	Region – Northern Ireland	0.941	0.067	
	Age – 18 – 24 years	0.797	0.072	
	Locale – Small Town (10 miles)	-1.250	0.096	