

METHODOLOGY

Survey Data and Population

This analysis uses the EU Loneliness Survey, a cross-sectional online survey conducted by the European Commission's Joint Research Centre (JRC) in November and December 2022. The survey collected data from 25,646 respondents aged 16 and older, spanning all 27 EU Member States, including a targeted subsample of approximately 1,000 from Ireland. Respondents were recruited through consumer panels (Cint platform). The survey instrument covered socio-demographics, municipality, migration, income, education, sexuality, technology use, and standardized loneliness scales (UCLA, De Jong Gierveld), alongside direct self-report loneliness measures and contextual variables relevant to urbanicity, digital engagement, and social connectedness (REF for The EU Loneliness Survey).

Data Preparation

Raw survey data were processed using Python and the pandas library. Key data cleaning and transformation steps included:

- Filtering for valid age (16–120 years), dropping out-of-range or missing demographic and thematic variables, e.g. dropping all blank responses and those with values of “Prefer not to say” or “Don’t know”.
- Recoding and grouping: age into categorical bands (16–24, 25–34, 35–44, 45–54, 55–64, 65+), income into Low (deciles 1–3), Medium (4–7), High (8–10), education into Primary or less, Secondary, and Post-secondary/Third Level, municipality type (rural, small/medium town, large town/city), municipality tenure into intervals, and sexuality into Heterosexual, Homosexual, Bisexual, and Other.
- Derivation of binary migrant flag: respondent and/or at least one parent born outside the current country.
- Technology engagement: recoded social media time/purpose/feelings into usage bands (Never, Light, Moderate, Heavy) and (Never, Seldom, Sometimes, Often), aggregating across multiple reported platforms/purposes.
- Construction of binary and summary loneliness measures: direct single-item loneliness (dichotomized: “some of the time” or more = 1, else 0), UCLA scale (sum of three items, higher = greater loneliness), and De Jong Gierveld (DJG) scale (recoded social and emotional items; higher = greater loneliness).
- Weighting: national and EU-level probability weights applied, using **w_country** for Ireland and **w_eu27** for all other countries, to yield estimates representative of the respective populations.

Analytical Themes and Variables

Analysis proceeded in six themes: Age, Migration, Municipality, Income & Education, Sexuality, and Technology. Each theme's analysis included both descriptive and inferential statistics, with comparative focus on Ireland versus the EU average.

Descriptive Analysis

- Weighted means and proportions for loneliness indices (direct, UCLA, DJG) by each theme and sub-group, separately for Ireland and the EU, using groupby-weighted aggregation.
- Tabulation of sample size, weighted N, and missingness for all outcomes and covariates to document data structure and exclusion criteria.

Inferential Analysis

Weighted Linear Regression (WLS)

For each theme, WLS models estimated the association between the primary independent variable (e.g., age band, migrant flag, municipality type, income/education group, sexual orientation, social media usage group) and loneliness outcomes, controlling for relevant covariates. All models included national probability weights. Covariates included age, income, education, municipality type, and migration as appropriate to the theme.

- For age: models estimated mean loneliness by age band, both within Ireland and EU, and included an interaction model (age band \times country group).
- For municipality: models included main effects for type, tenure, and interaction (type \times tenure); education and income were included as covariates.
- For migration: models included migrant flag, age band, education, income, and municipality type.
- For income and education: models included interaction (income group \times education group), age, municipality type, and post-hoc comparisons using Tukey's HSD where main effects were significant.
- For sexuality: models included sexual orientation group, age band, migrant flag, income group, and all relevant two-way interactions.
- For technology: models used highest social media usage group as the independent variable, controlling for age band, income group, municipality type, and included interaction models as appropriate.

Weighted Chi-square Tests

For all binary loneliness outcomes, weighted chi-square statistics (with expected and observed weighted cell counts) tested independence between key grouping variables (e.g., age band, migrant status, municipality type, income group, sexual orientation, social media usage) and loneliness prevalence, run separately for Ireland and EU.

ANOVA and Post-hoc Comparisons

For age and other categorical predictors, weighted ANOVA models assessed differences in mean loneliness scores across groups, followed by Tukey's HSD for pairwise group contrasts. Interaction effects were formally tested (e.g., age band \times country group), and all models included robust residual diagnostics.

Interaction Analyses

For each theme, key interactions (e.g., age × migration, income × education, sexual orientation × covariates, technology × municipality type) were modelled. Where interactions were significant, stratified group-level weighted means and comparative bar plots were generated.

Visualization

All figures and tables reflect weighted estimates. Bar plots, forest plots of regression coefficients, and histograms of model residuals were generated for each main and interaction analysis using seaborn and matplotlib. Forest plots display point estimates and confidence intervals for all modelled predictors, allowing direct comparison of effect sizes across groups.

Missing Data and Exclusions

Cases with missing data in critical variables for each analysis were excluded pairwise. A missingness table documented the proportion of excluded observations by group and variable.

Software and Reproducibility

All analyses were conducted using Python (pandas, statsmodels, scipy, seaborn, matplotlib) with scripts structured for reproducibility and export of all outputs (CSV, PNG, TXT). Weighting was applied at all analytic stages. Post-hoc outputs and residual diagnostics were reviewed to check model assumptions.