2024-2025

Development of Residential Energy Archetypes in Northern Ireland

Final Report

Building Energy Informatics Limited

Table of Contents

Project Overview

The following tables provide a summary of the project details.

Project Title	Development of Residential Energy Archetypes in Northern Ireland
Research Themes	Reduce Carbon Emissions / Decarbonisation

	Name	**Organisation**
Lead Applicant	Dr. James O'Donnell	Building Energy Informatics
Partner Applicant(s)	Dr. Usman Ali	Building Energy Informatics
Partner Applicant(s)	Sobia Bano	Building Energy Informatics

Executive Summary
The aim of this project is to evaluate the energy performance of residential buildings in Northern Ireland by developing representative residential archetypes that are based on Energy Performance Certificate (EPC) data. Using a structured approach, the study focuses on understanding current energy performance and identifying data-driven strategies for improvement. By synthesising findings from international best practices and national data, the project provides valuable insights to support informed decision-making, aligning with sustainability and climate targets.

ACRONYMS

ASHP Air Source Heat Pump

AWS Affordable Warmth Scheme

BER Building Energy Rating

BRS Boiler Replacement Scheme

DfE Department for the Economy (Northern Ireland)

DEAP Dwelling Energy Assessment Procedure

EESSH2 Energy Efficiency Standard for Social Housing (Version 2)

EHS English Housing Survey

EPC Energy Performance Certificate

GIS Geographic Information System

HVAC Heating, Ventilation and Air Conditioning

NEHM National Energy Modelling Framework

NI Northern Ireland

NIHCS Northern Ireland House Condition Survey

NIHE Northern Ireland Housing Executive

RdSAP Reduced Data Standard Assessment Procedure

ROI Republic of Ireland

SAP Standard Assessment Procedure

SEAI Sustainable Energy Authority of Ireland

SPSS Statistical Package for the Social Sciences

TABULA Typology Approach for Building Stock Energy Assessment (EU project)

UPRN Unique Property Reference Number

Background

A comprehensive understanding of residential building archetypes and typologies is essential for assessing and enhancing the energy performance of housing stocks. In the context of Northern Ireland, such an understanding can inform policies and initiatives aimed at economic development, particularly through energy efficiency improvements and retrofitting strategies. Classifying buildings into specific archetypes allows for targeted energy performance assessments. This facilitates the development of tailored retrofitting strategies to improve energy efficiency across different building types.

Understanding the distribution and characteristics of building archetypes enables policymakers to design effective regulations and incentives. This ensures that energy efficiency standards are appropriately applied across various building types, promoting sustainable development. A review of residential building archetypes emphasises their role in informing policy decisions related to building energy consumption.

Improving energy performance of residential buildings is a critical step towards achieving sustainability and decarbonisation goals. In Europe and the UK, several studies explored archetype or typology development to classify residential buildings based on shared characteristics.

One prominent example is the **TABULA project** (2009-2014)[1], which developed residential building typologies for several European countries, including the Republic of Ireland and England. These typologies classified buildings using parameters such as construction period, size, and energy-related features, offering a systematic framework to evaluate and enhance energy performance (Table 1).

In the **Republic of Ireland**, the 2014 TABULA typology[2] identified common residential building types and provided detailed energy profiles for each. This work facilitated targeted retrofit strategies and informed policy decisions, significantly contributing to the pathway for improving the energy efficiency of the Irish housing sector. The project identified 34 distinct building typologies, categorised across ten construction age bands ranging from pre-1900 to post-2010. These typologies are developed using the national Building Energy Rating (BER) database, the Dwelling Energy Assessment Procedure (DEAP) which is based on the UK SAP tool, and other sources such as census data and SEAI reports. The typologies accounted for variations in wall construction (e.g., solid stone, hollow block, and cavity walls), insulation levels, and heating systems.

Building on this, the Republic of Ireland's 2023 **National Energy Modelling Framework (NEMF)**[3] advanced the archetype approach by incorporating data from 680 individual heat demand archetypes. This model integrates both physical and consumer attributes, providing a nuanced description of energy demand across residential, service, industry, and agricultural sectors.

[1] TABLULA Project (2009-2014)
[2] Republic of Ireland TABULA typology (2014)
[3] National Energy Modelling Framework (NEMF, 2023)

Similarly, in **England**[4], the 2014 development of a national building typology enabled a comprehensive assessment of energy performance across diverse residential types. This approach informed the design and implementation of retrofit measures tailored to specific building categories, furthering national energy efficiency objectives. These typologies are derived from the English Housing Survey (EHS) and classified across eight construction age bands (pre-1919 to post-2010). The classification included dwelling types such as single-family houses, terraced houses, multifamily dwellings, and apartments (Table 2).

Table 1 Key summary of Republic of Ireland and England TABULA typology Development

Aspect	**Republic of Ireland**	**England**
Age Bands	10 distinct bands (e.g., pre-1900, 1900–1929, etc.).	8 distinct bands (e.g., pre-1919, 1919–1944, etc.).
Building Types	Detached, semi-detached, terraced houses, and apartments.	Single-family homes, terraced houses, multifamily dwellings, and apartments.
Properties Count	1 million	6,200 Properties (survey data only)
Data Sources	BER database, Irish Census, SEAI reports.	English Housing Survey (EHS).

Table 2 Summary of Data required for Ireland and England TABULA typology Development

Data Category	**Description for Both Countries (England and ROI)**
Physical Characteristics	Wall type (e.g., cavity, solid, stone), roof type, floor type, insulation levels.
Energy Systems	Heating systems (e.g., boilers, heat pumps), fuel types (e.g., gas, oil, electricity).
Key metrics	Primary Energy Demand (kWh/m²/year), Carbon Dioxide Emissions ($kgCO_2/(m^2 * year)$), and Energy Ratings (e.g., BER in the Republic of Ireland or EPC in England)

[4] England TABULA typology (2014)

Studies conducted in **Northern Ireland**[5] evaluated the energy performance of dwellings before and after improvement measures implemented under the **Affordable Warmth Scheme (AWS) and the Boiler Replacement Scheme (BRS).** These schemes provided grants to households to support energy efficiency improvements. The evaluation employed the UK Government's Standard Assessment Procedure (SAP) energy modelling methodology, specifically the Reduced Data SAP (RdSAP), which uses standardised assumptions to compare dwellings independent of occupant behaviour. One challenge in the study is incomplete data provided by the Northern Ireland Housing Executive (NIHE) for the pre-improvement state of dwellings. To address this, an inference procedure assigned each dwelling to an archetype based on characteristics such as type, age, and wall structure. These archetypes used data from the 2016 Northern Ireland House Condition Survey (NIHCS), which provided inputs for geometric dimensions and fabric heat loss values. This allowed for accurate SAP calculations even when direct measurements are unavailable (Table 3). BRE's SAP model simulated the effects of installed measures, allowing consistent energy, cost, and CO_2 savings calculations.

Table 3 Summary of data used for archetypes development in Affordable Warmth Scheme (AWS) and the Boiler Replacement Scheme (BRS) projects

Category	**Subcategory/Parameter**
Dwelling Types	Detached, Semi-Detached Mid-Terraced, End-Terraced, Detached Bungalow, Attached Bungalow, Ground Floor Flat, Mid-Floor Flat, Top-Floor Flat Unknown Archetype
Fabric Heat Loss Parameters	Wall U-value (W/m²K), Window U-value (W/m²K), Floor U-value (W/m²K), Roof U-value (W/m²K), Loft Insulation Thickness (mm)
Dimensions	Number of Storeys, Total Floor Area (m²), Ground Floor Area (m²), Roof Area (m²), Wall Area (m²), Party Wall Area (m²), Window Area (m²), External Door Area (m²) and Internal Door Area (m²)

Similarly, The **English Archetypes model**[6] serves as an established framework for large-scale domestic energy modelling across the UK (England). It is developed to provide a representative set of housing archetypes for analysing energy use, thermal comfort, and indoor air quality at a national level. The model is designed to facilitate building design optimisation and performance analysis by offering simplified but generic input parameters, including geometry, fabric, equipment, and operational characteristics. This model is derived from an extensive analysis of English housing stock, primarily using data from the **English**

[5] Affordable Warmth and Boiler Replacement Scheme Evaluation (2020-2022)
[6] English Archetypes Model

Housing Survey (EHS) 2010-11[7]. Key variables such as built form, floor area, construction age, and fabric characteristics are statistically analysed to develop a representative set of archetypes. Internal layouts are based on typical access patterns of dwellings from similar form, size, and age. Cluster analysis further refined the identification of groups with similar housing characteristics, leading to the development of a standardised set of English housing archetypes.

A recent study conducted in **Plymouth (2024)**[8] explores the feasibility of Air Source Heat Pump (ASHP) installations across five prevalent housing archetypes in the region. These included a Victorian bay-windowed terrace, crosswall flat-roofed terrace, Easiform house, bay-fronted 1930s bungalow, and a 1930s–1960s semi-detached house. The study assessed the viability of ASHP adoption by analysing the thermal performance, construction characteristics, and energy efficiency levels of these archetypes. The research provided key insights into the practical challenges and benefits of retrofitting heat pumps in different housing types by using archetype-based modelling. The findings emphasised the importance of tailored retrofit strategies to accommodate varying architectural designs and fabric efficiencies. This approach aligns with Northern Ireland's need for archetype-driven energy modelling, where region-specific housing characteristics must be considered for effective decarbonisation and sustainable heating solutions.

Another archetype-based study, conducted in 2023, focused on the Scottish social housing sector as part of the **Energy Efficiency Standard for Social Housing (EESSH2) Review**[9]. This research aimed to develop a set of 24 archetypes, of which 12 represent over 90% of the total number of dwellings to support housing associations and policymakers in achieving decarbonisation targets. By categorising social housing units based on their construction type, energy performance and suitability for retrofit measures, the study proposed a pattern book approach to guide the implementation of energy efficiency improvements and low-carbon heating systems.

The EESSH2 study provides a relevant comparison for Northern Ireland's housing sector, where social housing constitutes a significant portion of the residential stock. By adopting a similar archetype-based approach, stakeholders in Northern Ireland could streamline retrofit efforts, optimise funding allocation and develop effective energy efficiency policies tailored to the region's unique housing characteristics.

Although previous archetype-based studies provided valuable insights into energy modelling, retrofit feasibility and policy development, they are now considered outdated due to changes in the building stock, technologies and policy contexts. More importantly, these models lack the distinct residential characteristics of Northern Ireland. The NI housing stock differs in terms of construction methods, architectural styles, materials, and energy performance, meaning that archetypes developed elsewhere cannot be directly applied. As a result, there is a growing need to develop region-specific housing archetypes that capture the unique attributes of Northern Ireland's homes. Such an updated, data-driven model would provide policymakers, housing providers, and researchers with reliable tools to inform retrofit strategies, optimise funding decisions, and align energy efficiency improvements with local sustainability goals.

[7] Housing Survey (EHS) 2010-11
[8] Heat pump feasibility & housing archetypes in Plymouth (2024)
[9] Review of social housing archetypes to support EESSH2 review (2023)

Ultimately, a tailored archetype framework would support more effective and equitable decarbonisation efforts across the region's residential sector.

Objectives

- **Conduct a comprehensive literature review** to identify key insights, best practices, and data requirements for residential energy performance archetypes, culminating in a summary report and data specification document.
- **Develop and model residential building archetypes** specific to Northern Ireland, with detailed documentation of assumptions, inputs, and methodologies, delivered in a technical report.
- **Deliver final project outputs**, including a full report with guidelines, an Excel-based analysis tool, relevant coding and modelling frameworks and an online presentation to support dissemination, future research, and stakeholder engagement.

Methodology

The methodology for developing Residential Energy Archetypes in Northern Ireland consists of five key steps: data collection, pre-processing, segmentation, characterisation, and quantification (Figure 1). These steps form a systematic framework for understanding the energy performance and characteristics of the housing stock in the region.

Figure 1 Overarching methodology for Northern Ireland's building archetypes development.

Data collection

The first step, data collection, is critical to the entire process as it forms the foundation for the subsequent steps. This involves gathering comprehensive data on the residential building stock, focusing on both geometric and non-geometric information.

Data Requirements:

- **Geometric Data:** Information about the physical structure of the buildings, including:
  - Dwelling type (e.g., detached, semi-detached, terraced)
  - Building envelope details (e.g., walls, windows, roof, floors)
- **Non-Geometric Data:** Information related to building systems and energy performance, such as:
  - Envelope U-values
  - Construction assemblies (e.g., insulation types, materials)
  - Heating, Ventilation, and Air Conditioning (HVAC) systems and their efficiencies
  - Renewable energy features (e.g., solar panels, heat pumps)

Northern Ireland's Energy Performance Certificates (EPCs) provide a comprehensive dataset for this step, detailing energy ratings, insulation levels, heating system efficiency, and renewable energy adoption. EPC data are standardised and reflects typical occupancy and behaviour assumptions. However, EPCs present certain limitations: coverage is not universal across the housing stock, additionally, the data often lack granular detail on the building envelope (e.g. specific wall, floor, and roof constructions), which can constrain the accuracy of retrofit modelling and technical assessments. Other useful data sources that could help in this project include **Property Valuation Data (Northern Ireland)** which contains details on property attributes such as size, age, construction type, and market value, providing critical inputs for archetype classification, quantification, and economic analysis of retrofit strategies. Property valuation and spatial data supports quantification by helping to determine the distribution and count of each residential archetype within the overall building stock Similarly, **Pointer and GIS Data** includes spatial and location information to map building distributions and energy performances across regions, enabling regional archetype development and targeted retrofit planning.

Data Pre-Processing

Once the raw EPC and other useful data are collected, they need to be cleaned and structured to ensure consistency and accuracy using **data pre-processing** step. This process involves converting categorical energy efficiency ratings into numerical U-values, mapping construction year bands and removing duplicate property records based on the most recent assessment date. Building envelope categories as provided in energy efficiency ratings range from “Very Poor” to “Very Good”. To make these data usable for analysis as the categories are converted into approximate U-values, reflecting the thermal performance of building elements. Since the construction year has a significant impact on a building’s thermal properties so proper mapping into age bands is required for archetype development.

To prevent duplicate entries, the Unique Property Reference Number (UPRN) is used to uniquely identify individual buildings. Since some properties have multiple EPC assessments over time, only the most recent record is retained, as it reflects the latest energy efficiency improvements. For the purposes of this study, developing representative archetypes for Northern Ireland, the most recent EPC per property is used. If national-scale energy modelling is undertaken based on the generated archetypes, the year of the latest EPC would be used to represent the temporal context of the dataset. Duplicate UPRNs are identified, sorted by assessment date, and filtered to retain only the most recent EPC entry for each property.

Furthermore, missing or inconsistent data must be addressed. Any anomalies, such as unrealistic values or incorrect formatting, are corrected to maintain data integrity. The dataset now becomes structured, accurate, and ready for further analysis in energy performance modelling by performing these steps.

Segmentation

The second step, **segmentation** which focuses on dividing the collected data into meaningful groups that represent different building archetypes. This step determines the number and types of archetypes needed to effectively characterise the residential building stock. Segmentation criteria are selected based on their relevance to energy performance. In this project, two key criteria are used: *dwelling type* and *construction year*. *Dwelling type* refers to the physical configuration of the buildings, such as detached houses, semi-detached houses, terraced houses, apartments, or bungalows. This classification is essential as energy use and heating requirements vary significantly based on topology. For example, detached houses typically have higher energy demands due to their larger surface area exposed to the external environment, compared to terraced houses which benefit from shared walls. *Construction year*, the second criterion, is another critical factor influencing energy performance. Older buildings often consume more energy due to outdated construction methods and materials, while newer buildings generally benefit from advancements in energy-efficient technologies and stricter building regulations. Segmenting the housing stock by construction year (e.g., pre-1919, 1920–1944, 1945–1979, 1980–1999, post-2000) captures these variations and helps group buildings with similar energy characteristics.

Characterisation

The third step, **characterisation**, involves describing the detailed physical, thermal, and operational properties of each archetype. This step defines how each archetype behaves in

terms of energy performance and helps create a representative model of the housing stock. Characterisation parameters include the thermal properties of the building envelope, such as insulation levels, U-values, and airtightness; construction materials used in walls, roofs, windows, and floors; and the types of HVAC systems, including their efficiency and fuel sources. A data-driven approach is used to define the most typical values for each characterisation archetype. For categorical variables, the mode is selected to represent the most frequently occurring value, such as common wall or heating types. For numerical variables, the median is used to avoid distortion from outliers and better reflect typical conditions, such as floor area or U-values. This method ensures that each archetype is based on realistic and representative building characteristics found in the dataset.

Quantification

The fourth step, **quantification**, determines the prevalence of each archetype within the building stock. This involves calculating the number of buildings in each archetype category and assigning a "weighting coefficient" to each archetype based on its representation in the population. The quantification process relies on national building statistics or census data to aggregate the results, which may include property valuation records and spatial data to ensure accurate representation of the building stock across different regions and typologies. This information is critical for scaling the impact of energy efficiency policies or retrofit strategies, as it highlights which archetypes dominate the housing stock and therefore have the greatest potential for energy savings.

The methodology's systematic framework ensures a comprehensive and reliable approach to developing residential energy archetypes. Data collection lays the groundwork by providing a rich dataset, while segmentation ensures that the diversity of the housing stock is captured. Characterisation adds depth by defining the unique properties of each archetype, and quantification provides a clear understanding of their prevalence. Validation further strengthens the framework by ensuring the developed archetypes reflect real-world conditions accurately. Together, these steps provide a robust foundation for energy modelling, retrofit strategies, and sustainable development initiatives tailored to NI's residential sector.

Contributing datasets

To achieve the project's objectives, we used several datasets for analysis and archetype development.

The **Northern Ireland Energy Performance Certificate (EPC) Dataset** is the primary data source for archetype development and provides critical insights into the energy efficiency of residential properties across Northern Ireland[10]. EPCs assign energy ratings to properties, ranging from **A (most efficient)** to **G (least efficient)**, and include recommendations for improvements to reduce energy use and carbon emissions. Below is a summary of key statistics and how this data supports archetype development using the January 2024 EPC dataset.

[10] EPC data originates from the Northern Ireland register which is managed by the Department of Finance NI.

Key Statistics from the EPC Dataset

1. **Energy Ratings Distribution**:
    - **Most Common Ratings**: The majority of dwellings fall within the **C (39%)** and **D (41%)** rating bands.
    - **Poor Performance**: Approximately **10%** of properties have an energy rating of **E, F, or G**, highlighting the need for significant energy efficiency improvements.
    - **Highly Efficient Properties**: Only **5%** of properties are rated **A or B**, representing newer and well-insulated buildings.
2. **Dwelling Types and Efficiency**:
    - **Detached Houses**: Generally less efficient, with a significant proportion in the **D and E bands** due to larger surface areas and older construction.
    - **Flats and Apartments**: More commonly rated **C or above**, likely due to shared walls and modern construction practices.
    - **Terraced and Semi-Detached Houses**: Fall between detached houses and flats, with many rated **C or D**.
3. **Heating Systems**:
    - Over **60%** of homes use oil-fired central heating, which is less efficient compared to gas or renewable systems.
    - Only **10%** of homes have modern, renewable heating solutions like heat pumps or solar thermal systems.
4. **Insulation Levels**:
    - Properties constructed before **1980** tend to lack adequate wall and roof insulation, contributing to higher energy use.
    - Post-2000 properties generally meet modern insulation standards, with many achieving **B or C ratings**.
5. **Geographical Variations**:
    - Urban areas tend to have higher energy efficiency ratings due to newer construction and access to gas heating.
    - Rural areas, where oil heating is prevalent and homes are often older, show lower ratings on average.

The 2024 **EPC dataset** contains a total of **643,258 records**, offering valuable insights into building types across Northern Ireland. The breakdown of building types is as follows: Semi-Detached properties account for 29.8% (188,671 records), Mid-Terrace properties make up 26.9% (170,653 records), Detached properties represent 25% (158,262 records), and End-Terrace properties constitute 15.8% (99,901 records). Additionally, Enclosed Mid-Terrace and Enclosed End-Terrace properties account for only 1.2% each, with 7,853 and 7,658 records, respectively. A remaining 10,260 records (1.6%) fall into other or unclassified categories.

While this dataset provides a solid foundation for archetype development, several key issues need to be addressed. One major challenge is the absence of Unique Property Reference Numbers (UPRNs) for approximately **554,101 records, which accounts for 86.2%** of the dataset. This poses a significant challenge in deduplicating EPC records and ensuring an accurate representation of the building stock. Another issue is that all building envelope parameters are currently in categorical format, requiring conversion into U-values to facilitate energy performance assessment.

Moreover, a considerable portion of the dataset contains missing values, which must be addressed before archetype classification. Given the scope and duration of the project, a

practical solution is to use pre-processed data from previous studies in 2021 to streamline the development process while ensuring data reliability and consistency. These issues have been addressed by applying existing DfE preprocessing methods previously developed for the 2021 EPC dataset, allowing for improved data consistency and usability. The 2021 dataset therefore provided sufficient information to address these gaps without replacing the core January 2024 dataset. Older entries are the issue and only needed to be addressed once.

Archetype Development Results

Summary of Pre-processing

The initial data pre-processing phase is successfully completed using EPC data from Northern Ireland, resulting in the following key figures:

- **2024 EPC Dataset:** 643,258 records
- **2021 EPC Dataset:** 570,420 records

All collected EPC data underwent comprehensive cleaning and structuring procedures. This included the conversion of categorical EPC ratings into numerical U-values for critical building envelope components—walls, roofs, and floors—using existing SPSS scripts established in previous studies. Construction years are mapped systematically into defined age bands to reflect distinct energy performance characteristics. Duplicate EPC entries for the same properties are identified using the Unique Property Reference Number (UPRN). To avoid duplicate entries, the Unique Property Reference Number (UPRN) is used to identify individual buildings. Many properties have multiple EPCs due to improvements over time. For this analysis, only the most recent EPC is kept, as it best reflects the current energy performance. Duplicate UPRNs are sorted by assessment date, and only the latest entry is retained.

Furthermore, during the data pre-processing stage, building types were aligned to ensure consistency across the dataset. One challenge identified early was that apartments were not classified under the 'Building Type' field, as they were only indicated under 'PROPERTY_TYPE' as flats. This discrepancy created a gap in the structural classification of the housing stock. To address this—particularly important for archetype modeling where complete and consistent structural categories are required—all properties marked as flats were reclassified as apartments. Other property types such as mid-terrace, end-terrace, and enclosed terraces were grouped under the broader category of terraced homes, while semi-detached and detached homes were classified accordingly. This reclassification ensured that no property type was overlooked due to inconsistencies between different fields in the source data. This approach ensured a complete and coherent categorisation of the housing stock. As a result, terraced homes accounted for 32.9%, semi-detached homes 27.0%, detached homes 25.6%, and apartments 14.5% of the dataset.

Key Insights and Findings

1. Segmentation Variables

Detailed analysis identified the following effective segmentation variables for archetype development:

- **Building Type:** A key factor due to significant variations in energy use patterns among detached, semi-detached, terraced houses and apartments
- **Construction Year (Age Band):** Highly influential in defining thermal performance and construction characteristics.
- **Energy Rating:** Directly relevant for identifying opportunities for energy-efficiency retrofits and improvements.
- **Heating fuel type**: The **main heating fuel type** is also evaluated as a potential segmentation variable. However, since approximately 89% of properties use oil and gas, and only a small fraction use electricity or other fuel sources

2. Archetype Segmentation Proposals

Five proposed approaches to segmentation are developed based on identified variables. These approaches are presented to a DfE steering group and are outlined below:

- **Segmentation approach I (Building Type and Age Band):**
    - **44 archetypes** created based on combinations of dwelling type and construction year bands.
    - Advantage: Provides clear segmentation for easy quantification using property-enhanced datasets, enabling accurate representation of stock variations.
    - **Disadvantage:** May not directly reflect thermal performance or energy efficiency without additional attributes (e.g., energy ratings or fuel type).
- **Segmentation approach II (Building Type and Energy Rating):**
    - **28 archetypes** identified by categorising buildings according to dwelling type and EPC energy ratings.
    - Advantage: Facilitates targeted identification of properties suitable for retrofitting by focusing on lower energy-rated segments.
    - **Disadvantage:** Requires weighted-average quantification due to incomplete EPC stock coverage, posing some analytical limitations.
- **Segmentation approach III (Building Type and Main Fuel):**
    - **16 archetypes** segmented by dwelling type and primary fuel source (e.g., oil, gas, electricity).
    - Advantage: Offers insights into the influence of fuel types on energy efficiency.
    - Disadvantage: Limited variability and the dominance of certain fuels (e.g., gas) reduce its effectiveness as a standalone segmentation framework.
- **Segmentation approach IV (Building Type, Main Fuel and Age Band):**

  A total of **176 archetypes** are created by segmenting dwellings based on:
    - **Building Type**: Detached, Semi-Detached, Terrace, Apartment
    - **Main Fuel**: Primary heating source (e.g. gas, oil, electricity, etc.)
    - **Age Band**: Construction period band
    - Advantage: This approach allows exploration of how primary fuel types interact with building age and form, providing insights into energy efficiency potential.
    - Disadvantage: However, due to the limited variability in some fuel categories and dominance of specific types (e.g. oil & gas), its utility as the primary segmentation framework is limited.
- **Segmentation approach V (Building Type, Energy Rating and Age Band):**

  This segmentation produced **216 archetypes** by combining:
    - **Building Type**: Detached, Semi-Detached, Terrace, Apartment
    - **Energy Rating** (originally A-G, then grouped into AB, C, D, E, and FG for practical coverage)
    - **Age Band**: Construction period band
    - Advantage: Integrates structural and thermal performance characteristics, so highly relevant for retrofit strategy, policy planning and scenario modelling.
    - Disadvantage: Some loss of granularity due to grouped ratings and potential bias from EPC data, particularly in underrepresented property types or regions.

The DfE steering group agreed to use Segmentation Approaches IV and V.

3. Archetype Characterisation

Initial characterisation analysis identified **54 out of 94 features** as particularly suitable for detailed archetype modelling due to their completeness and consistency across EPC records. Key characterisation parameters include:

- Thermal properties of the building envelope (walls, floors, roofs): U-values.
- Insulation types and levels.
- Heating, ventilation, and air-conditioning (HVAC) system characteristics and efficiencies.
- Renewable energy system installations (e.g., solar panels, heat pumps).

For each proposed archetype segmentation, characterisation is performed using statistical measures[11]:

- **Mean** values are calculated for continuous parameters (e.g., U-values, heating system efficiency).
- **Mode** is applied to categorical parameters (e.g., insulation type, construction materials).

These statistical approaches ensure that each archetype is accurately and representatively characterised by the most typical energy performance features present in the residential building stock.

4. Quantification Approach

Following the development of residential archetypes using different segmentation approaches, the next critical task is **quantification** — determining how many dwellings in Northern Ireland (NI) fall into each archetype category. This quantification process is essential for scaling the findings and for guiding retrofit policy, funding allocation, and technical interventions with real-world relevance.

Why Quantification Is Needed

While the segmentation produced a diverse and representative set of archetypes based on building type, construction age, and heating fuel (or energy rating), the **raw counts derived from the working dataset** could not be directly interpreted as representative of the entire housing stock in NI. The dataset is a sample that, while rich in features, may have over- or under-represented certain building types or categories due to collection method, availability of energy performance data, or other factors.

To ensure that the final archetype outputs would meaningfully reflect the actual NI housing landscape, a **calibration process** is necessary — one that would align the sample-based archetype data with authoritative, full-population housing statistics.

[11] Mean values were used for continuous variables to provide a representative average across each archetype group. For categorical parameters, the mode (most frequently occurring value) was applied to reflect the most common characteristic within each group, ensuring realistic and interpretable archetype profiles.

Archetype Count from Sample Data

Initially, the working dataset (which included building features such as type, age band, and main fuel) is grouped by the three key segmentation variables:

- **Building type** (Detached, Semi-Detached, Terrace, Apartment)
- **Construction age band**
- **Main fuel type**

This allowed for the creation of individual archetypes and provided a count of how many sample records belonged to each. This gave a **relative distribution of archetypes** within the dataset. However, this distribution is reflective only of the sample and not of the entire NI housing stock.

Obtaining Baseline Housing Stock Distribution

To calibrate the archetype counts, authoritative housing stock figures from the **Northern Ireland Property Valuation dataset** are used. This dataset provided the actual number of dwellings in Northern Ireland by building type and served as the baseline, or ground truth, for the calibration process.

For example, it is known that approximately 289,631 dwellings in NI are detached houses, while 208,219 are semi-detached, and so on. This provided the **true distribution by building type**, but not by age band or fuel type, which are only available in the EPC dataset (Figure 2).

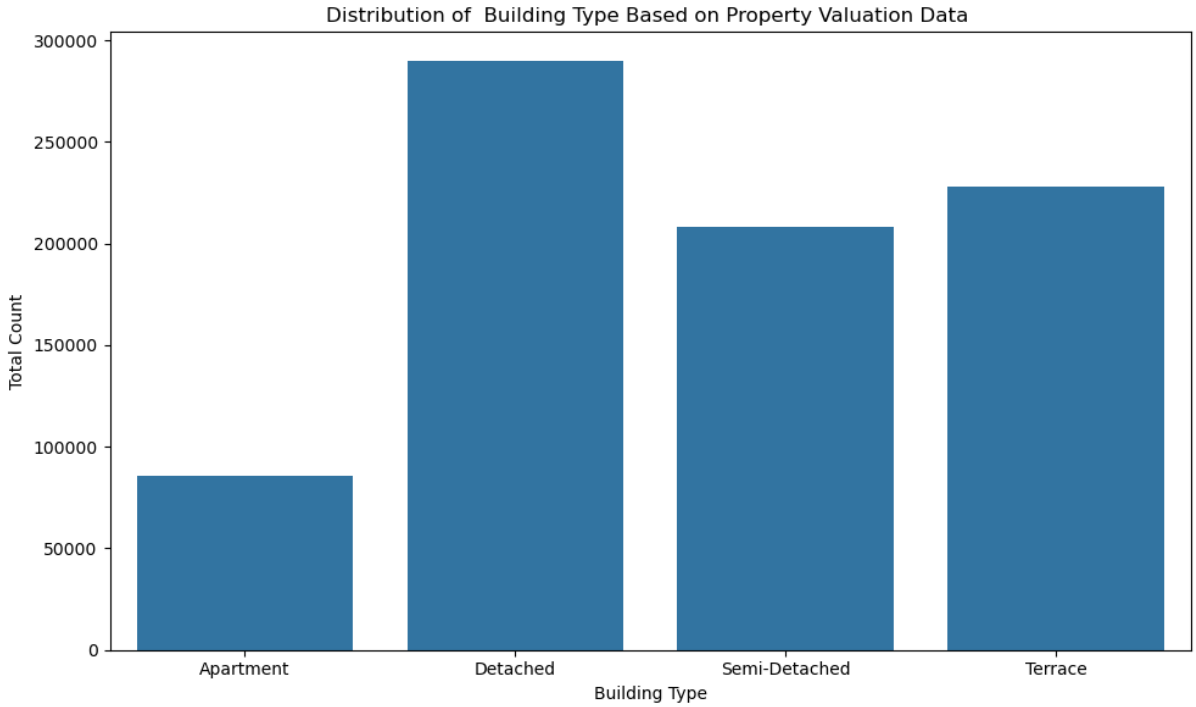

Figure 2 Total Residential Building Stock Distribution by Building Type in Northern Ireland

Proportional Scaling Within Each Building Type

As the **total building type counts** are only known at the NI level (not the full cross-tabulated archetypes), a proportional approach is applied. The share of each archetype **within its own building type** (e.g., the percentage of detached homes that are pre-1945 and oil-heated) is calculated using the sample data.

Then, these proportions are applied to the real-world total count of that building type in NI. For instance, if 10% of detached homes in the sample are built before 1945 and used oil, that same 10% is applied to the 289,631 real detached homes to estimate the **true number of homes in NI matching that archetype**.

This method preserved the **relative structure** of archetypes from the sample while anchoring the total count to verified housing data, effectively correcting for any sampling bias.

National-Level Calibration and Weighting

Once each archetype's adjusted count is estimated, it is possible to:

- Derive **calibrated percentages** showing the share of each archetype across the total NI housing stock.
- Assign **weighting coefficients** to each archetype, which can later be used to scale modelled outputs (e.g., energy demand, retrofit potential) to the full population.

This step is critical for ensuring that simulation, cost-benefit analysis, and policy modelling reflect actual conditions in NI, rather than being based on raw, potentially skewed, sample distributions.

The outcomes from using the Segmentation Approaches IV and V are now detailed:

- **IV) Building Type, Main Fuel, and Age Band**
- **V) Building Type, Energy Rating, and Age Band**

Quantification Summary – Building Type, Main Fuel, and Age Band (IV)

This segmentation approach combined three key variables to develop residential archetypes:

- **Building Type**: Detached, Semi-Detached, Terrace, Apartment
- **Main Fuel** : Gas, Oil, Electricity, Others
- **Age Band**

A total of **176 archetypes** are generated and quantified using proportional calibration, aligning sample proportions with verified housing stock totals across Northern Ireland.

Key Findings:

- **Fuel Type Dominance**:
    - **Oil** remains the most prevalent heating fuel across all building types, particularly in **Detached** and **Semi-Detached** homes constructed between **1950–2006**.
    - **Gas** is widely used in **Terraced houses** and **Apartments**, especially in **post-1990 constructions**.
    - **Electricity** usage is higher in **Apartments**, especially those built **after 1985**, and represents a modest share in newer **Detached** and **Terraced homes**.
    - **Other fuels** (e.g., solid fuel or unspecified systems) account for a small but consistent share across all types.
- **Building Type Trends**:
    - **Detached homes** show a heavy reliance on **oil heating**, particularly in older and rural stock.
    - **Apartments** demonstrate greater adoption of **gas and electricity**, especially in newer buildings.

    - **Semi-Detached and Terraced homes** show a balanced mix of oil and gas, with a transition toward gas in newer dwellings.
- **Temporal Shifts**:
    - **Electric heating** remains present but secondary, with higher presence in **Apartments** and **social housing**.
    - The pre-1973 stock that makes up the 20% calibrated coverage across all types is still heavily reliant on oil, indicating high retrofit potential.

Quantification Methodology:

- Sample data is grouped by building type, fuel type, and age band.
- Property Valuation data totals (by building type) are used to **scale sample proportions**, correcting for sampling bias.
- **Calibrated counts and percentages** are calculated to reflect actual distribution across NI's housing sector.

Challenges and Limitations:

- **Fuel type dominance**: Over 89% of homes use **oil or gas**, limiting segmentation value from smaller fuel categories (electricity, others).
- **Limited data granularity**: Limited cross-tabulations statistics of age and fuel, requiring proportional estimation.
- **Small sample size** in some archetypes (e.g., electricity in pre-1930 Detached homes 0.14%), introduces potential uncertainty in calibrated figures.

This segmentation approach provides valuable **conclusion** insights into the interaction between **fuel type**, **building form**, and **construction age**. While **fuel diversity is limited**, the quantification highlights clear **temporal and typological patterns** that can guide retrofit targeting and low-carbon heating strategies across Northern Ireland's residential sector.

Quantification Results Summary – Building Type, Energy Rating, and Age Band (V)

This quantification exercise is carried out using the segmentation framework that combines **Building Type**, **Energy Rating and Age Band**. The approach produced a total of **220 residential archetypes**, enabling a granular representation of Northern Ireland's housing stock for energy efficiency analysis and retrofit planning.

Key Findings:

- **High-prevalence archetypes included:**
    - **1950–1973 terraced dwellings:** 6.68%, primarily with a D energy rating
    - **1992–2006 detached dwellings:** 4.88% combined, heavily concentrated in the D band
        - 1992–1999: 2.19%, 2000–2006: 2.69%
    - **1950–1973 semi-detached dwellings:** 2.95%, mostly rated D
    - **Post-2000 apartments:** 3.68%, dominated by C-rated units
        - 2000–2006: 1.75%, 2007–2013: 0.84%, 2014 onwards: 1.09%
- **Energy performance patterns showed:**
    - **A and B-rated dwellings (AB)** are primarily found in **recent construction (post-2007),** particularly among **Detached, Semi-Detached, and Apartment properties.** These account for 8.08% of the stock:
        - Detached: 3.77%

    - Semi-Detached: 4.34%
    - Apartments: 1.01%
- **E and FG-rated dwellings** are still predominantly located in **older housing stock,** especially **Terraced and Detached homes built before 1978, which represent 15.34%** of the dataset. A significant portion (5.27%) was built **before 1919.**

Quantification Methodology:

- The working dataset is first grouped by **building type**, **construction age band**, and **energy rating category**.
- Using **official housing stock totals** by dwelling type (from property valuation sources[12]), a **proportional scaling method** is applied.
- Sample-based proportions within each dwelling type are used to estimate actual archetype counts across Northern Ireland. Within each building type, the percentage of properties falling into each archetype in the EPC sample is calculated. These proportions are then applied to the total number of that building type from the property valuation data. For instance, if 15% of detached homes in the EPC dataset fall into a certain archetype, that 15% is applied to the total number of detached homes in Northern Ireland to estimate how many homes of that archetype actually exist.
- This enabled the creation of **calibrated counts and percentages** that better reflect the real-world distribution of homes in each archetype category. This calibrated dataset is essential for reliable modelling, retrofit targeting, and policy planning, as it accounts for under- or over-representation in the original EPC data.

Key Challenges and Limitations:

- The EPC dataset, while rich in detail, is **not fully representative** of the entire NI housing stock. As a result, certain archetypes may be over- or under-represented within the sample, requiring calibration against official housing stock data to ensure accurate quantification archetypes
- **Low sample counts** are observed in rare archetypes, such as **AB-rated pre-1930 dwellings**, introducing higher levels of uncertainty in those estimates.

Outcome and Value:

- A set of **calibrated archetype counts** is produced for all 216 combinations.
- Each archetype is assigned a **weighting coefficient**, enabling its use in future modelling, simulation, or cost-benefit analysis.
- This quantification provides a critical foundation for evaluating the **retrofit potential**, **carbon savings**, and **energy demand** across the Northern Ireland residential sector.

[12] This is based on Crown Copyright and is reproduced with the permission of Land & Property Services under delegated authority from the Keeper of Public Records, © Crown copyright and database right (2025) NIMA MOU577.3.

Conclusion

The **Development of Residential Energy Archetypes in Northern Ireland** has provided a structured, data-driven framework for understanding the diverse and evolving characteristics of the region's residential building stock. Through detailed segmentation, characterisation, and quantification of over 600,000 EPC records, the study has identified representative building archetypes that reflect the real-world variation in energy performance, construction age, heating systems, and building forms across Northern Ireland.

Two advanced segmentation strategies such as **(IV) Building Type, Main Fuel, and Age Band** and **(V) Building Type, Energy Rating, and Age Band** are adopted to capture both physical and thermal dimensions of residential buildings. These approaches successfully generated 176 and 216 archetypes respectively, each calibrated using national housing statistics to ensure representativeness and practical relevance.

The findings reveal that a significant portion of dwellings, especially detached and terraced houses built before 1973, remain heavily reliant on oil heating and have poor energy efficiency ratings (D to G). These homes present the greatest opportunities for retrofit and decarbonisation interventions. In contrast, newer properties, particularly apartments and semi-detached homes constructed post-2007, show better energy performance (most commonly AB rating) and more frequent adoption of gas or electric heating systems. The majority of all dwellings fall into the D and E energy rating bands, underscoring the broad potential for improvement across the stock.

The resulting archetypes provide a robust foundation for energy modelling, retrofit planning, and targeted policy development. They offer a practical tool for assessing energy demand, estimating retrofit costs and benefits, and designing interventions that align with real housing conditions. The study highlights the need for ongoing data updates to keep archetypes current. This would include updated EPC datasets, property upgrades, records of retrofitting activities (e.g., insulation, heating system upgrades), and any changes in building stock characteristics or age band over time, the integration of emerging technologies such as heat pumps and solar PV, and the use of spatial mapping for regional planning. These updates are essential to ensure the model reflects the latest building conditions and retrofit potential. In conclusion, this project supports evidence-based decision-making and equips Northern Ireland with the analytical tools needed to meet its climate and energy efficiency goals through targeted and equitable housing upgrades.

Appendix

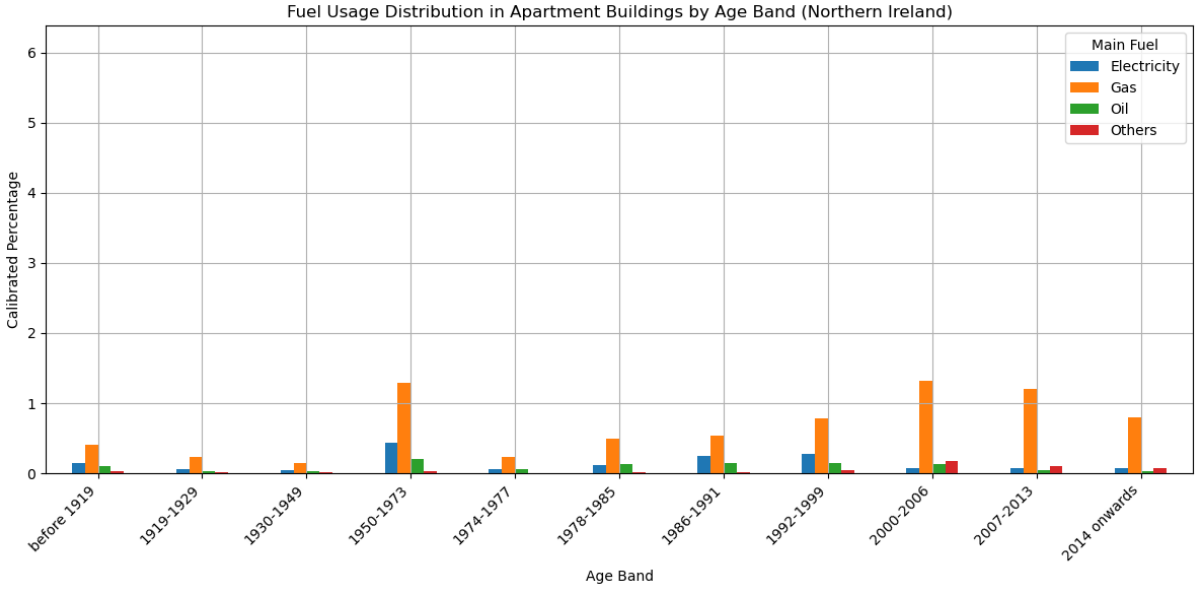

Figure 3 Fuel Usage Distribution in Apartment Buildings by Age Band (Northern Ireland)

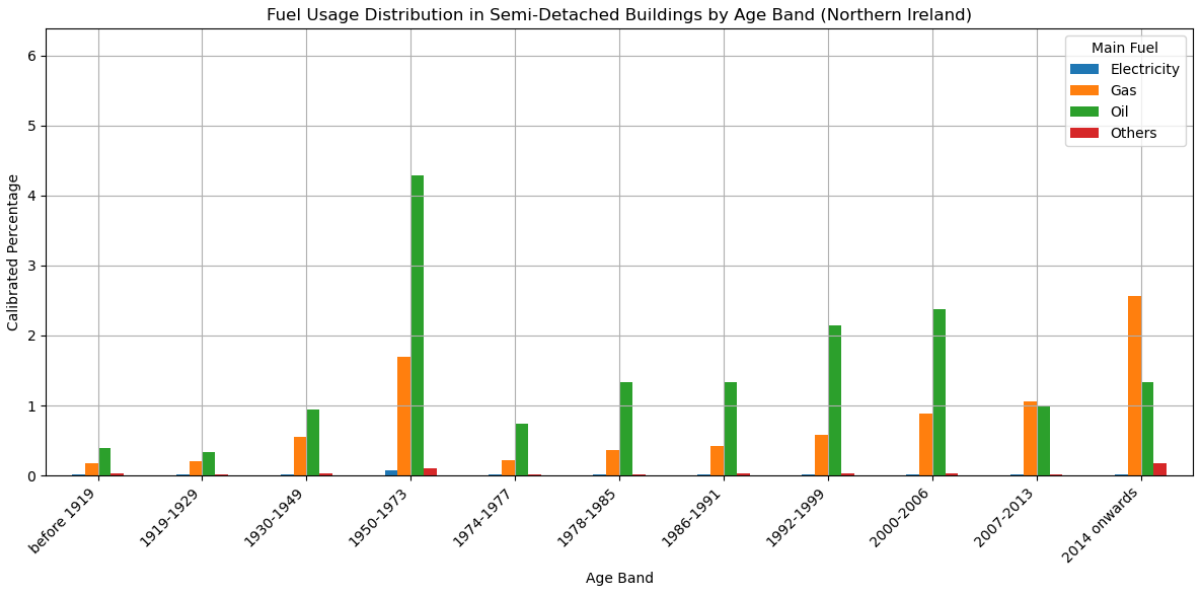

Figure 4 Fuel Usage Distribution in Semi-Detached Houses by Age Band (Northern Ireland)

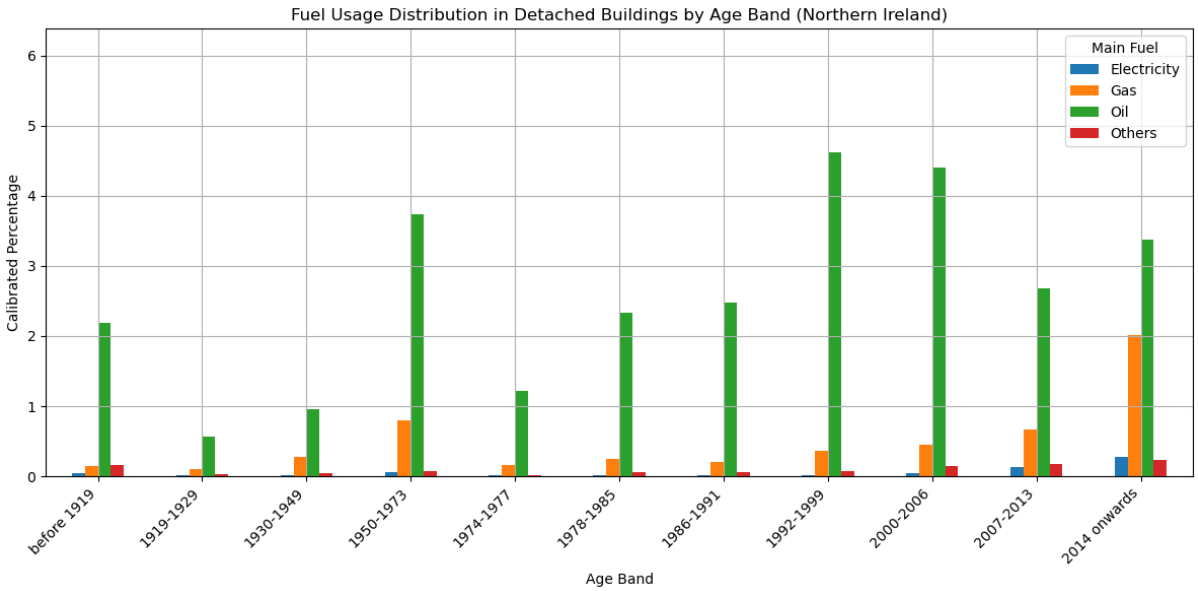

Figure 5 Fuel Usage Distribution in Detached Houses by Age Band (Northern Ireland)

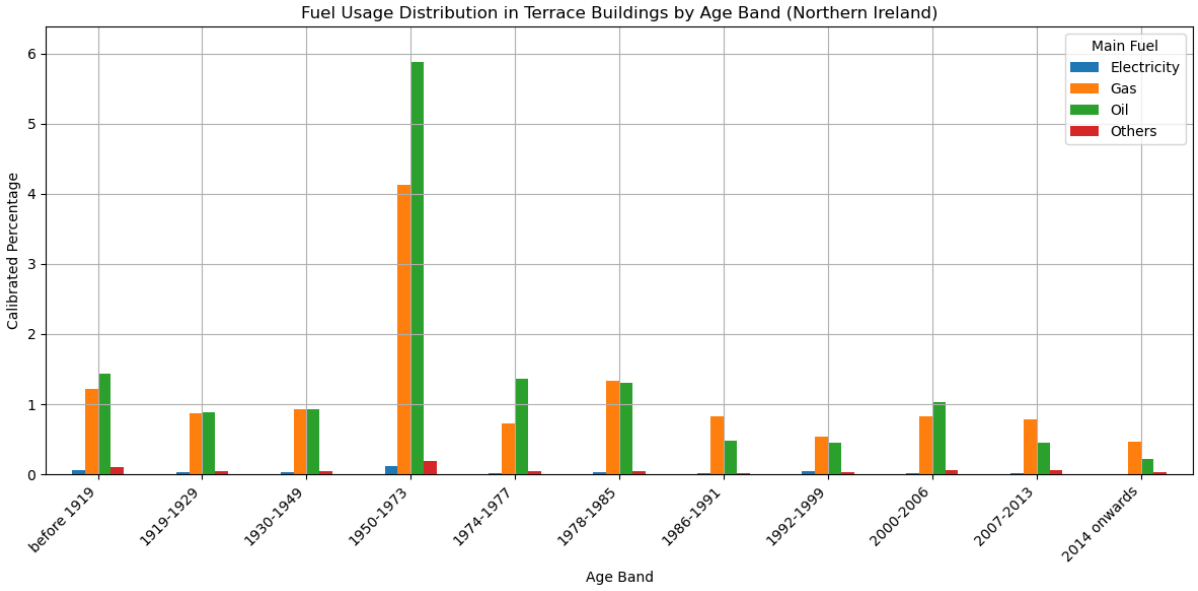

Figure 6 Fuel Usage Distribution in Terraced Houses by Age Band (Northern Ireland)

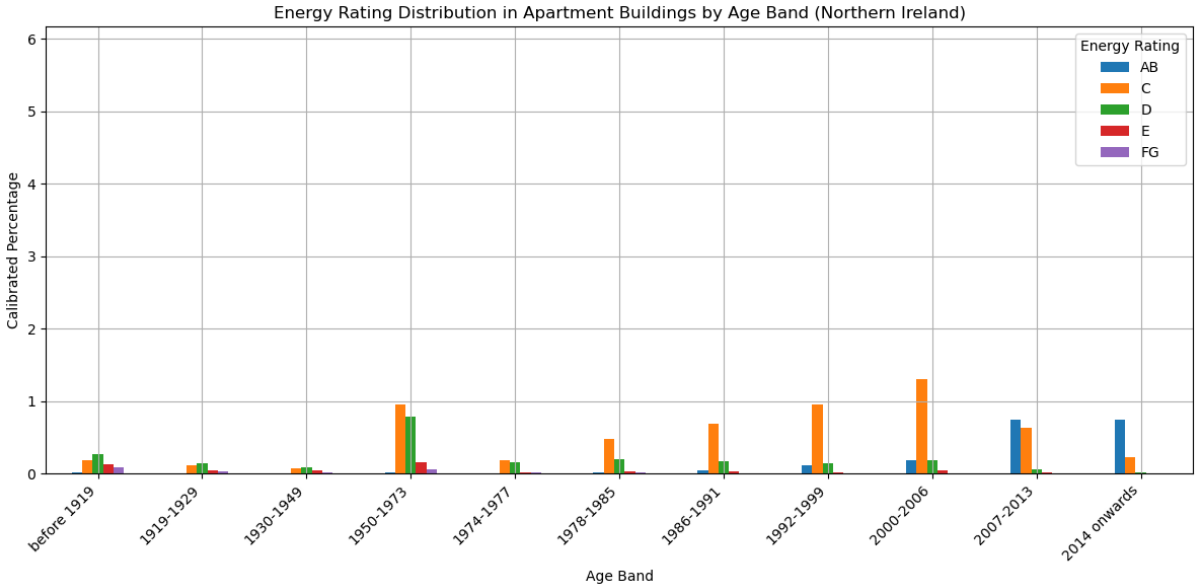

Figure 7 Fuel Usage Distribution in Apartment Buildings by Age Band (Northern Ireland)

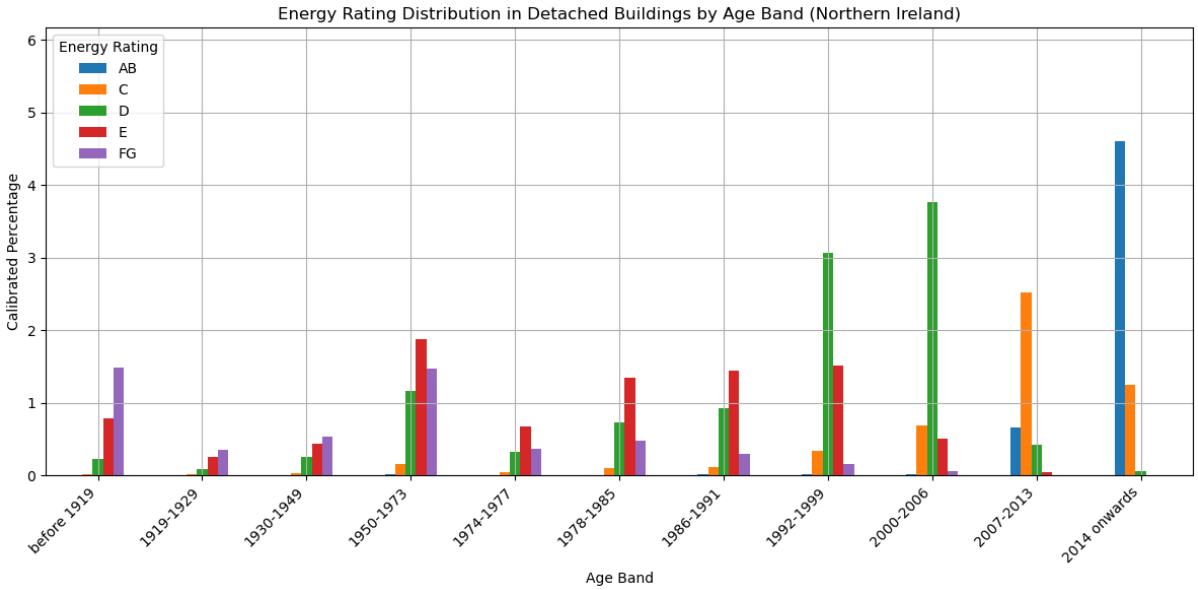

Figure 8 Energy Rating Distribution in Detached Buildings by Age Band (Northern Ireland)

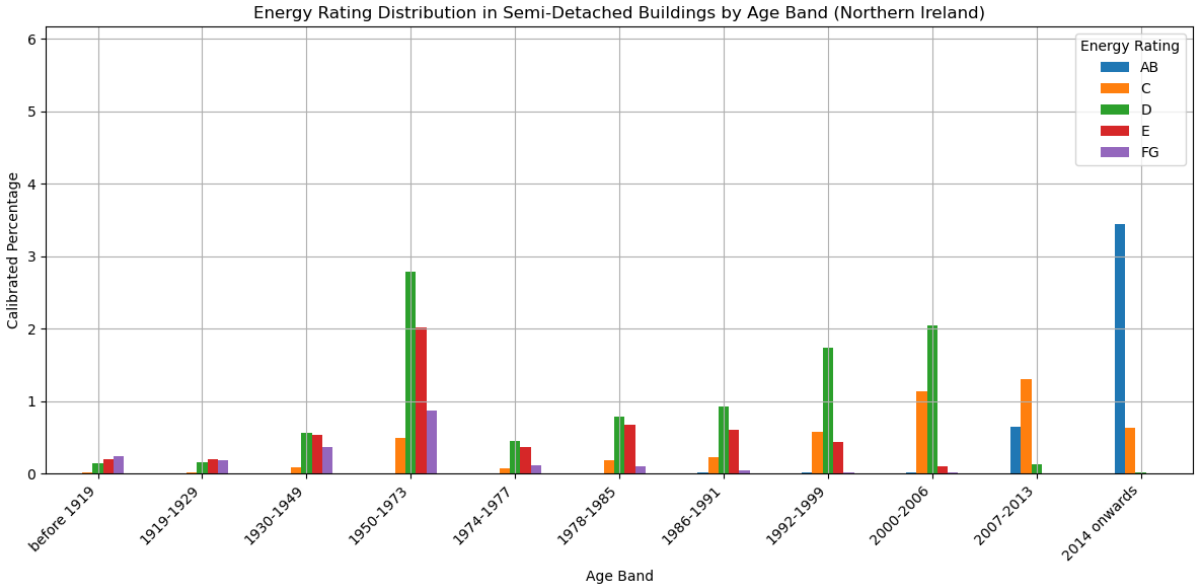

Figure 9 Energy Rating Distribution in Semi-Detached Buildings by Age Band (Northern Ireland)

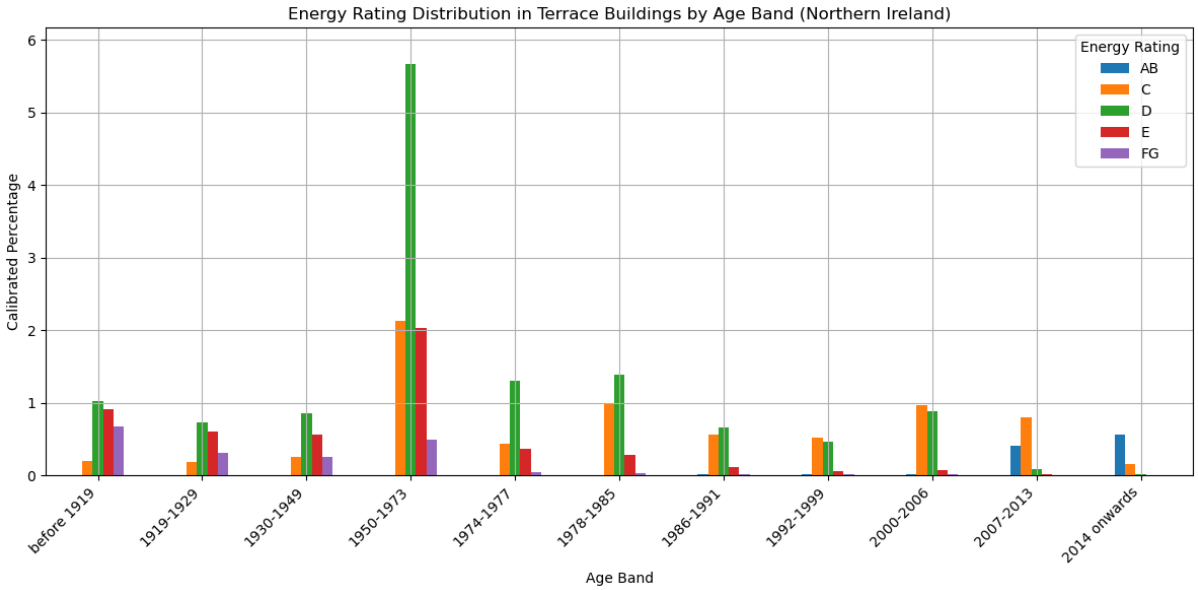

Figure 10 Energy Rating Distribution in Terrace Buildings by Age Band (Northern Ireland)