Census Project Report

Pascal Ugonna Akano
202445626
University of Hull
30th April 2025

Census Project **3**
Overview of the Dataset 3
Handling The Missing Values 3
Age 4
Marital Status 4
Gender 4
First Name 4
Infirmity 4
Religion 5
Relationship to Head of House 5
Demographic analysis **5**
Marital Status 7
Relationship to Head of House 8
Religion 9
Occupation Analysis **12**
Unemployment 12
Retired 13
Commuters 14
Regional Analysis **14**
House Occupancy 14
Population Density 15
Birth and Death Rates : A Comparative Analysis **16**
Recommendation and Suggestions **16**
Investment Priorities 16
What Should Be Built on the Unoccupied Plot? 16
Reference List **18**

Census Project

This report analyses census data from a fictional town situated between two major cities, connected by motorways. Modelled after the 1881 UK Census, the dataset supports urban planning decisions by identifying the most suitable development for an unoccupied plot, whether residential, healthcare, educational, or other public use.

The analysis evaluates demographic trends, commuting patterns, religious affiliation, healthcare needs, employment, and education, with final recommendations drawn directly from the data.

Data Cleaning

Overview of the Dataset

The dataset consisted of eleven columns, including demographic, social, and housing details. Although mostly complete, some fields had missing values, inconsistent entries, and minor errors. Cleaning involved standardising labels, correcting typos, and using logical rules to infer missing data where applicable.

```
census_data_copy.info() # Data overview
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8485 entries, 0 to 8484
Data columns (total 11 columns):
 #   Column                        Non-Null Count  Dtype 
---  ------                        --------------  ----- 
 0   House Number                  8485 non-null   int64 
 1   Street                        8485 non-null   object
 2   First Name                    8485 non-null   object
 3   Surname                       8485 non-null   object
 4   Age                           8485 non-null   object
 5   Relationship to Head of House 7836 non-null   object
 6   Marital Status                6420 non-null   object
 7   Gender                        8485 non-null   object
 8   Occupation                    8485 non-null   object
 9   Infirmity                     71 non-null     object
 10  Religion                      3638 non-null   object
dtypes: int64(1), object(10)
memory usage: 729.3+ KB
```

The dataset includes 11 columns, 8,485 entries, and missing or misformatted values in several fields, including Age and Relationship to Head of House.

Handling The Missing Values

Missing values are a common issue in census data, often caused by input errors or non-responses. Addressing them is critical to ensure accurate and unbiased analysis. In this dataset, fields like Infirmity, Marital Status, Religion, and Relationship to Head of House contained null entries that required careful cleaning before analysis could begin.

All data processing was performed using Python with the pandas library. The objective was to ensure that the dataset was both complete and logically coherent.

Age

The Age column initially contained string values and non-numeric entries, causing the data type to default to object. To correct this, the column was converted to numeric format using pd.to_numeric() with coercion enabled to handle invalid entries. Any resulting missing values were then imputed by calculating the median age of individuals within the same “Relationship to Head of House” group. For example, if an individual’s relationship was “Son”, their missing age was filled using the median age of all sons in the dataset. If the relationship group was unavailable or unclear, the overall median age was used as a fallback. Finally, the column was converted to int64 format for consistency.

Marital Status

In the Marital Status column, abbreviations like “S”, “M”, “W”, and “D” were replaced with their full forms. Empty entries were converted to NaN, and all such cases were linked to individuals under 18, who were recategorized as “Single (Underage)”, forming a new valid label. One remaining NaN was resolved using a related row with matching House Number, Street, and Surname.

Gender

The Gender column contained a mixture of formats such as "M", "F", "male", and "female". These were standardized using regular expressions and converted to either "Male" or "Female". There were no missing values in this column after cleaning.

First Name

The Gender column contained a mixture of formats such as "M", "F", "male", and "female". These were standardized using regular expressions and converted to either "Male" or "Female". There were no missing values in this column after cleaning.

Infirmity

The Infirmity column had a high number of missing or blank entries. In line with WHO classification practices and NHS population health assumptions, individuals without a reported health issue were assumed to be free of major impairments. Missing values were therefore imputed as "Healthy". This approach reflects guidance from the World Health Organization (WHO, 2020), which states that population surveys often rely on self-reporting and that absence of disclosure is generally interpreted as absence of condition — unless risk factors are known.

Religion

The Religion column contained a mixture of valid religious labels and unconventional entries such as “Jedi”, “Nope” and “Private”. These were replaced with "None", aligning with census

best practices where non-religious or undefined responses are grouped as secular or "No religion". For example, the UK Office for National Statistics (2021) grouped similar categories into “No religion” during national census reporting.

To fill in missing values, religion was forward-filled within households (grouped by Street, House Number, and Surname) under the assumption of religious consistency within families. According to Pew Research Center (2020), **91% of adolescents aged 13–17** report sharing some or all religious beliefs with their parents, supporting the assumption of household-level religious consistency.

Relationship to Head of House

This column was the most complex to clean due to typographical errors and missing values. Corrections included fixing entries like "Neice" to "Niece" and treating blanks as missing. Logical rules were applied: children under 18 sharing the head’s surname were classified as “Son” or “Daughter”, while unrelated adults were labeled as “Lodger” (if employed) or “Visitor” (if unemployed). Children with the same surname as a lodger were also considered lodgers.

In households without a designated Head, the oldest resident was assigned the role, aligning with UK Office for National Statistics (ONS) guidelines. Following these steps, all missing values were resolved, and household structures were made consistent.

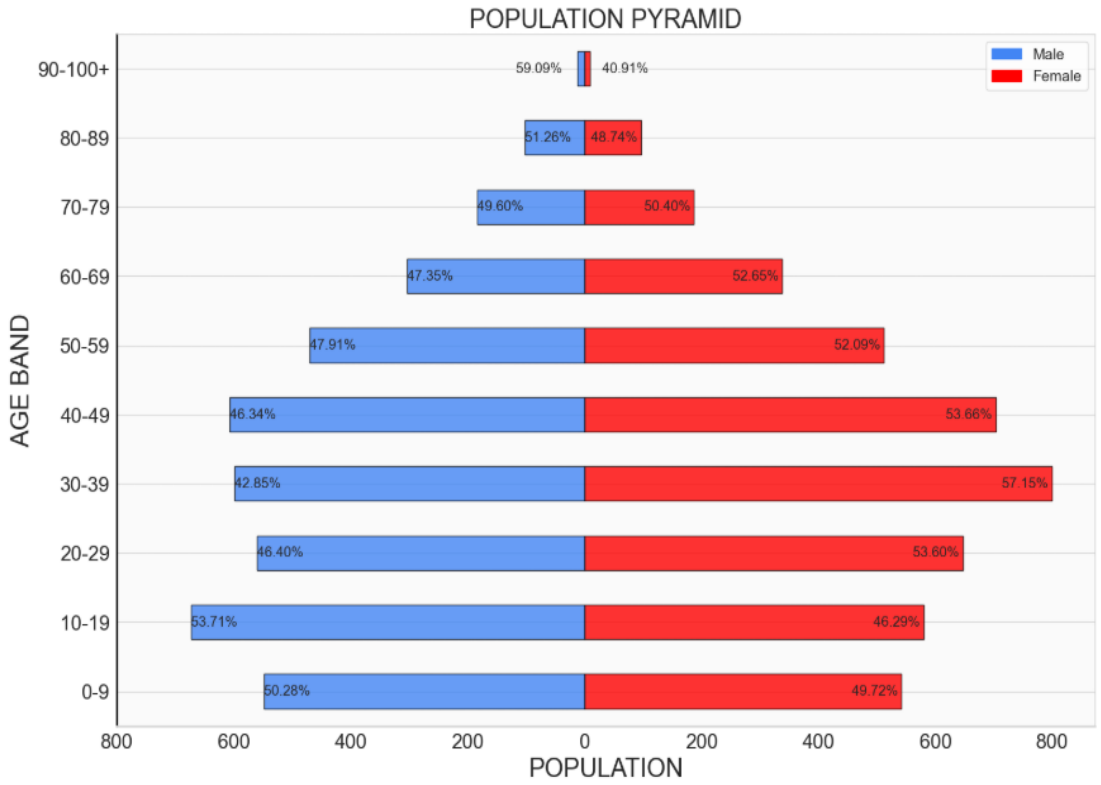
Other columns were complete and free of outliers. Full cleaning steps are detailed in the accompanying Jupyter Notebook.

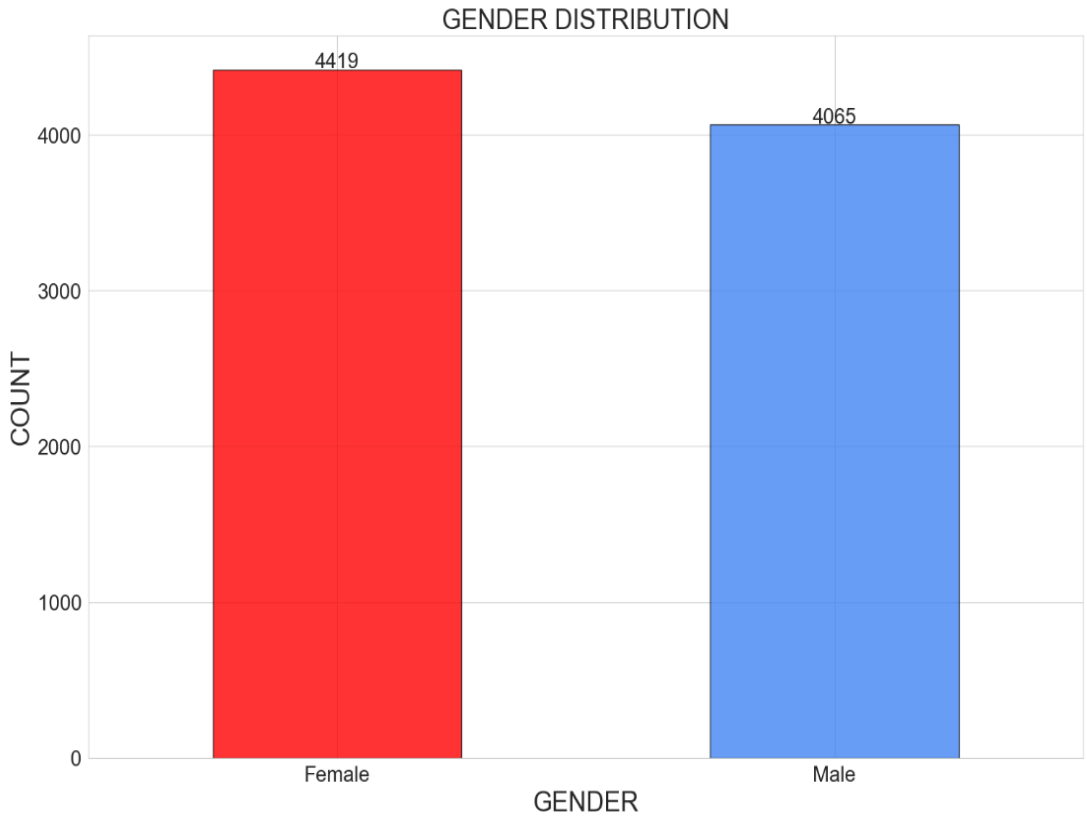
Demographic analysis

Demographic analysis helps identify the town’s social structure, health needs, and future growth trends. This section examines the distributions of marital status, age, gender, infirmity, the relationship to head of house, and religion to inform infrastructure planning.

The town’s age distribution displays a relatively wide base, suggesting a youthful and gradually expanding population. The average age is approximately 35, while the maximum recorded age is 105. The population declines steadily from age 60 onwards, reflecting expected life expectancy trends.

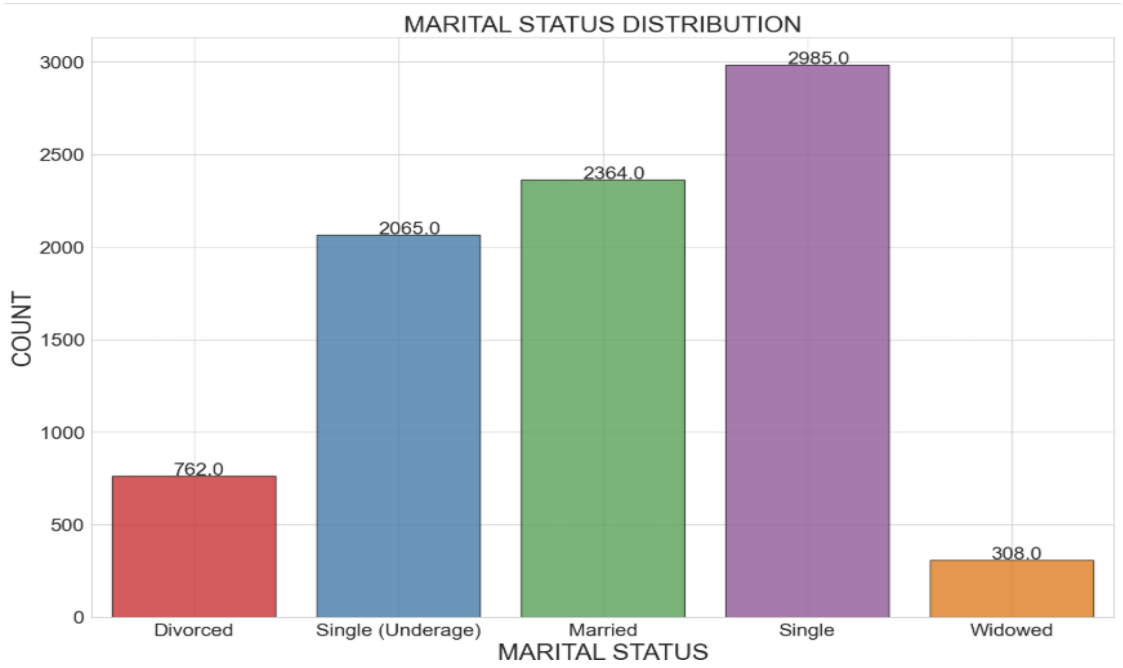
Age and Gender Distribution



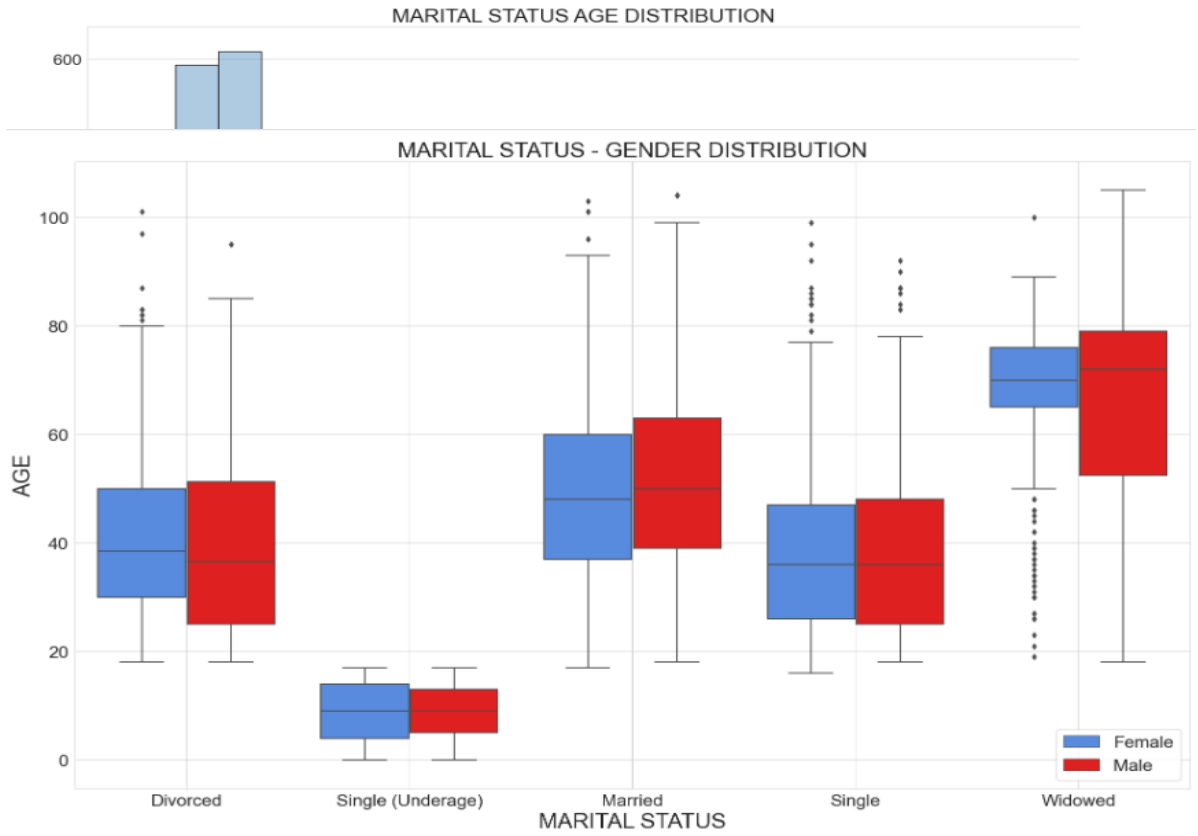


Gender-wise, 47.91% of the population are male, while 52.09% are female. The slight predominance of females becomes more evident in the older age brackets, likely due to higher female life expectancy. This gender distribution impacts the number of female-led households, which is explored further below.

Marital Status



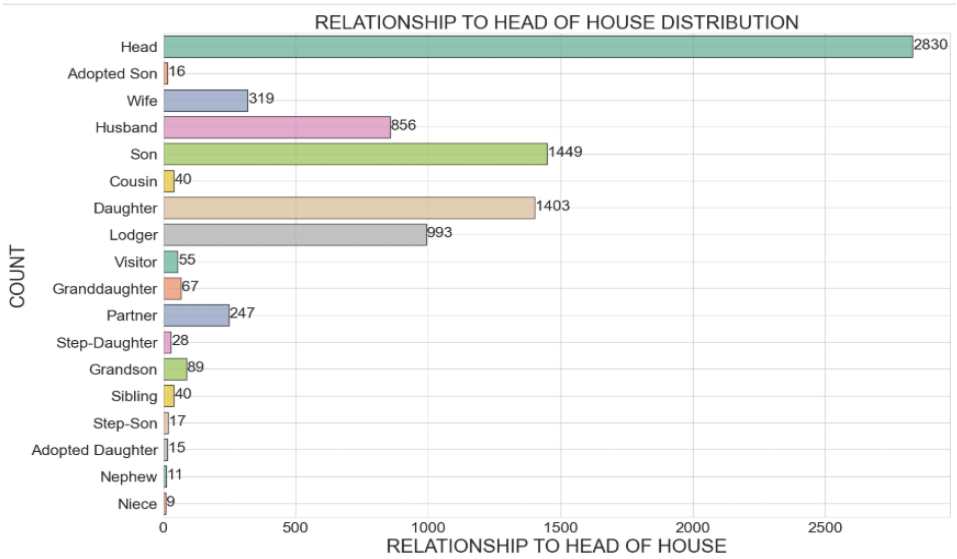
The visualisation shows the distribution of marital statuses, with Single being the most common and Widowed the least.



According to calculations from the accompanying Jupyter Notebook, 35.18% of adults are single, 27.86% are married, 8.98% are divorced, and 3.63% are widowed, while 24.34% of the population are classified as Single (Underage).The highest proportion of married individuals occurs at age 49, while single individuals are most common at age 20. Among the married, 49.7% are male and 50.3% are female. The marriage rate is 278 per 1000 people, the divorce rate is 90 per 1000, resulting in a marriage-to-divorce ratio of approximately 3:1.

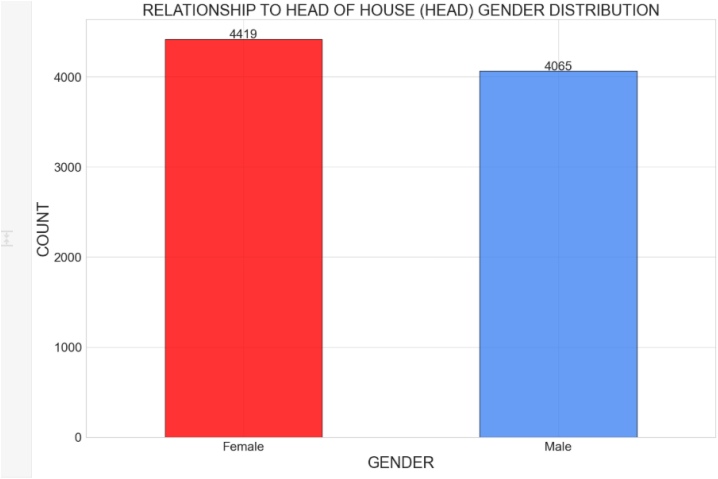
Relationship to Head of House

Household roles in the town show both traditional family structures and social diversity. The most frequently recorded roles were “Head”, “Son”, “Daughter”, and “Wife”. These highlight a population anchored in nuclear family systems. However, the dataset also includes non-traditional roles such as “Lodger”, “Visitor”, and “Partner”, reflecting mixed housing situations.



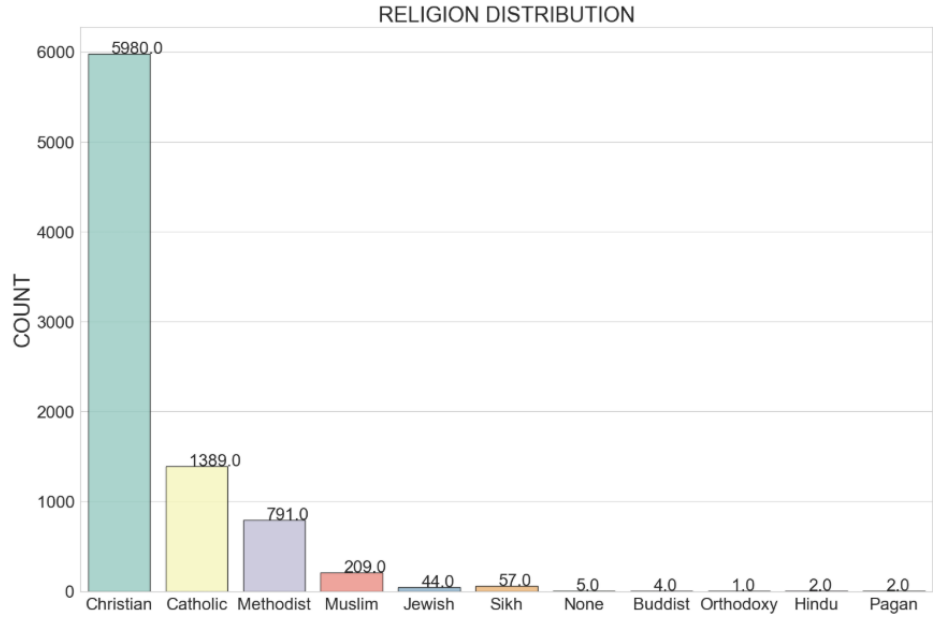
Lodgers and visitors were typically adults over 18, unrelated to the household head, and had different surnames. Working adults were classified as Lodgers, while unemployed individuals became Visitors, indicating the presence of students, low-income renters, and temporary residents.

Children sharing a surname with a lodger were also labeled Lodger, suggesting alternative family arrangements, such as minors under the care of siblings or relatives. In households without a designated head, the oldest resident was assigned that role, often revealing multi-generational living structures.

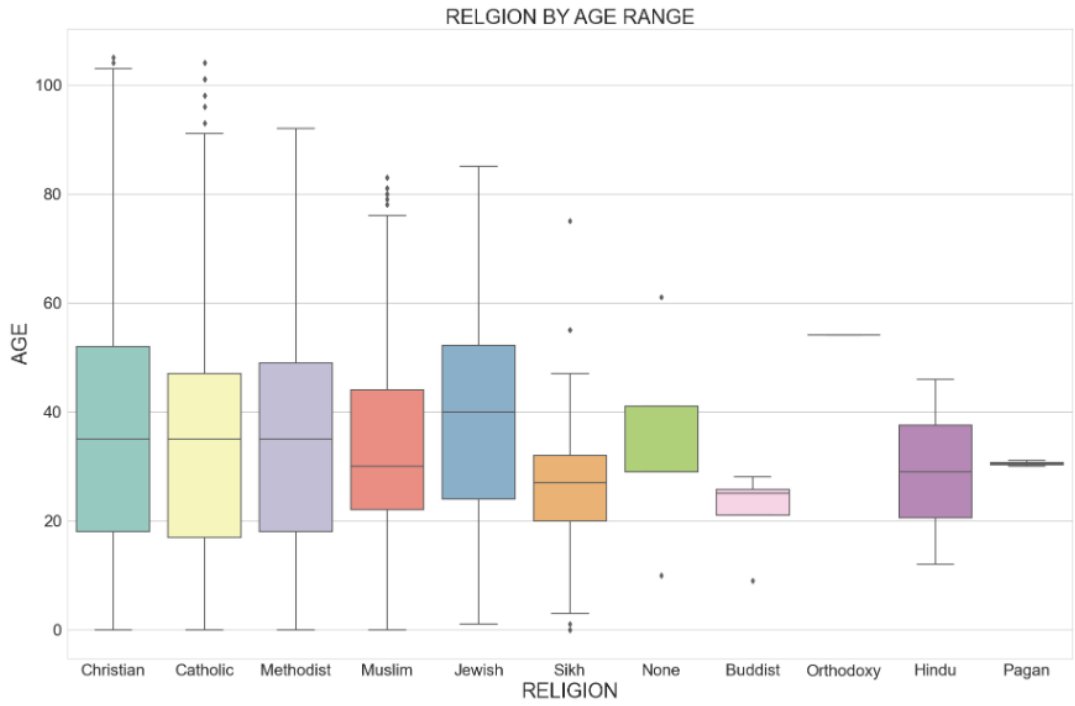


Notably, females (4,419) outnumbered males (4,065) as heads of house, reflecting patterns of single motherhood, widowhood, or increasing female-led households.Together, these findings illustrate how the town balances traditional family structures with modern living realities. They also point to policy areas such as affordable housing, youth welfare, and women-focused support systems.

Religion

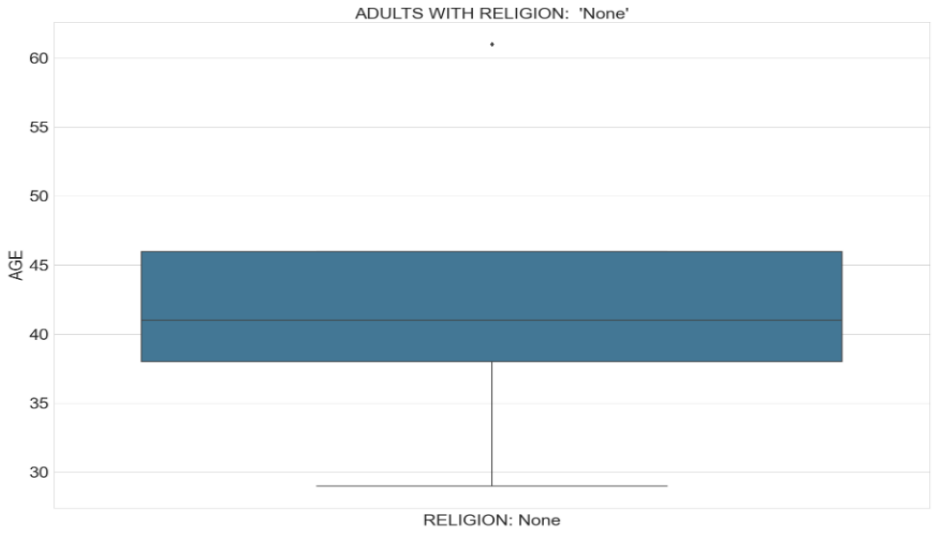


From the visualisation above, Christianity is clearly the dominant religion in the town, with 5,980 individuals, followed by Catholic (1,389) and Methodist (791). Religions such as Muslim, Jewish, and Sikh are present in smaller numbers. Only 5 people identified with "None", indicating that secularism is not common in this community.



Although national trends show a rise in non-religious identification, the low count of "None" in this dataset may be due to cultural norms of the time and the cleaning method used — where missing values were filled with the most frequent religion (Christianity).

The age range for different religions also varies. Christians and Methodists show a wide spread across all age groups, while smaller religions like Sikh and Hindu tend to have younger populations. Adults with “None” had a median age of 41 and an average age of 36, suggesting that non-religious identity may occur later in life in this community.

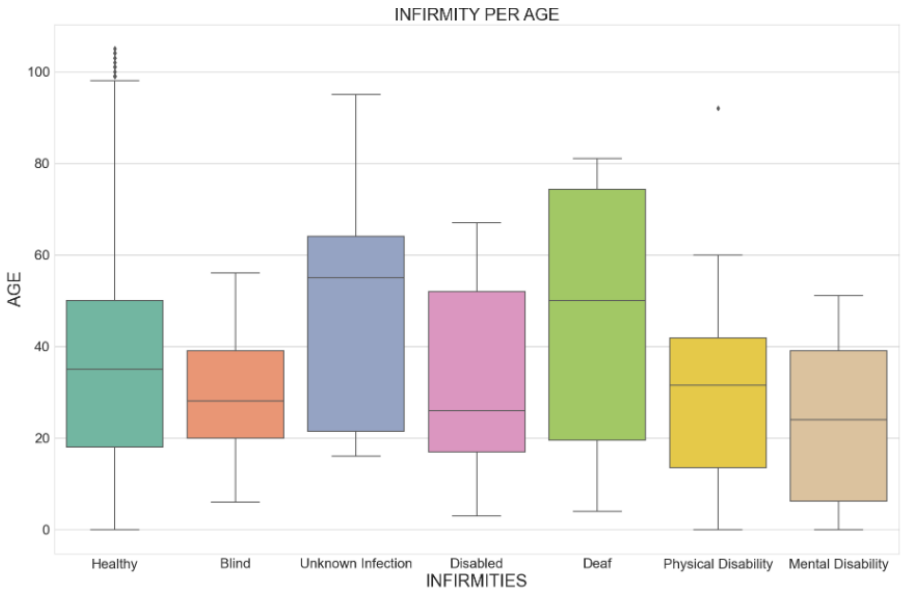


Religion was transmitted from parents to children in approximately 1.0% of households, with 2,925 out of 2,932 families showing consistent religious affiliation.

Infirmity

]:

	0	1	2	3	4	5	6
Infirmity	Healthy	Physical Disability	Unknown Infection	Deaf	Mental Disability	Blind	Disabled
Counts	8422	14	10	10	10	9	9



The visualisations show that 99.26% of the population report being healthy, including those aged 65 and above. This older group maintains high self-reported health levels, suggesting a well-supported community (NHS England, no date).

The average age of healthy individuals is 35, consistent with the town's overall mean. Cases of physical disability, mental illness, and deafness are rare and spread across different age groups, as shown in the infirmity boxplot.

Given these findings, building a healthcare facility is not a priority. Instead, development should focus on services that support the town’s predominantly healthy and active population.

Occupation Analysis

This occupation analysis provides insight into employment patterns, unemployment rates, retired individuals, and commuter profiles.The table above shows the top 10 occupations. Students make up 19.65% of the population, followed by University Students (6.38%), Unemployed individuals (6.46%), Children (5.86%), and PhD Students (0.2%). These figures

	0	1	2	3	4	5	6	7	8	9
Occupation	Student	Unemployed	University Student	Child	Phd Student	Landscape Architect	Theatre Director	Lobbyist	Librarian, Academic	Communications Engineer
Count	1667	548	541	497	17	16	15	15	14	14

highlight a strong youth presence, suggesting that education and part-time work are major lifestyle factors in the town.

Unemployment

The town has a total of 548 unemployed individuals. Age and gender breakdowns show a higher percentage of females than males are unemployed. According to the OECD definition, the labour force includes individuals aged 15 to 64, and based on your dataset, this gives an unemployment rate of approximately 8.62%, which is high compared to the UK's 2023 average of 3.7% Census Project Report. This suggests the town may benefit from retraining or upskilling programs, especially for unemployed adults under 30.

Retired

Although “Retired” is not among the top 10 occupations, a separate analysis of entries labeled as such shows that retirees make up approximately 8.04% of the population. The average age of a retired individual is 77, with a minimum of 68 years. Gender distribution is nearly balanced: 50.36% female, 49.64% male.

This points to a stable, aging segment that may require tailored care facilities or age-friendly infrastructure in the near future.

Commuters

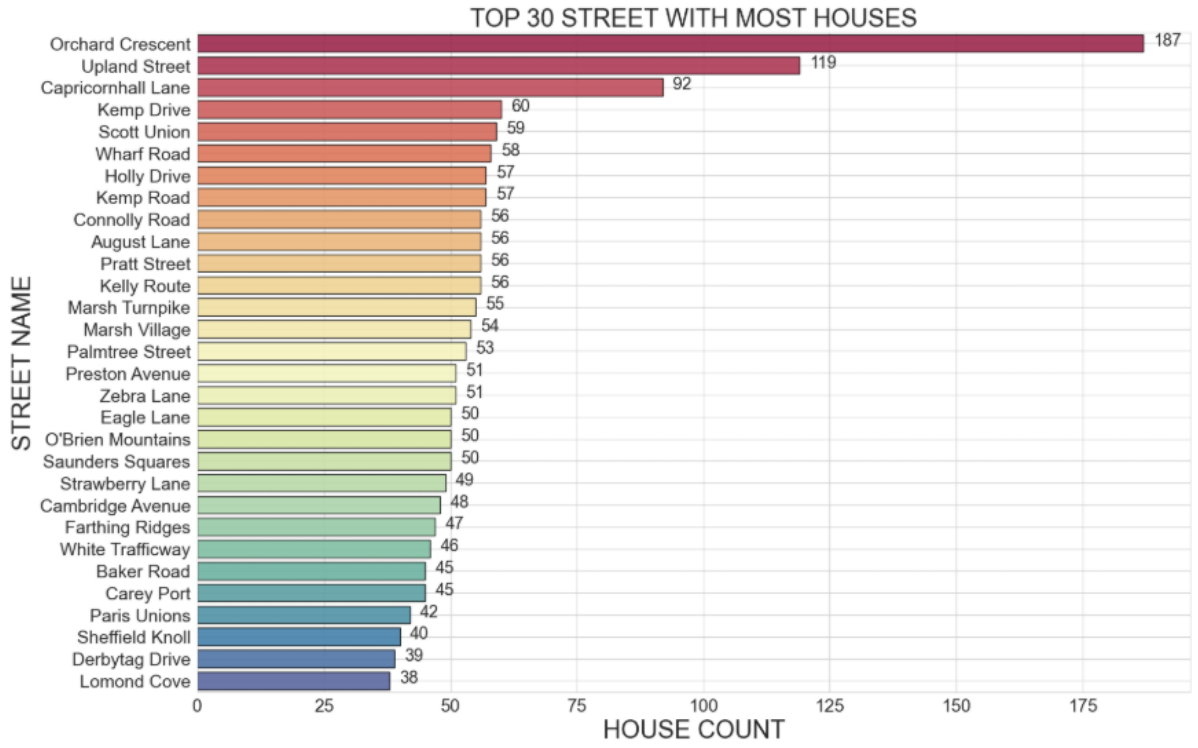
Given that the town lacks a university, University Students, PhD Students, Visitors, and Lodgers are classified as commuters. When combined, these groups represent 16.4 percent of the population . Since they reside in the town but likely travel frequently, the need for improved transportation options, such as bus lines or shuttle services, becomes a viable consideration. However, this commuter size alone may not justify building a train station, unless population trends or demand increases significantly.

Regional Analysis

This section focuses on house occupancy and population density to identify areas of high residential concentration and assess whether housing development or decongestant policies should be considered.

House Occupancy

According to the dataset, the town consists of 104 unique streets and 2,820 households, based on combinations of Street and House Number. The average number of occupants per house is approximately 3 individuals.

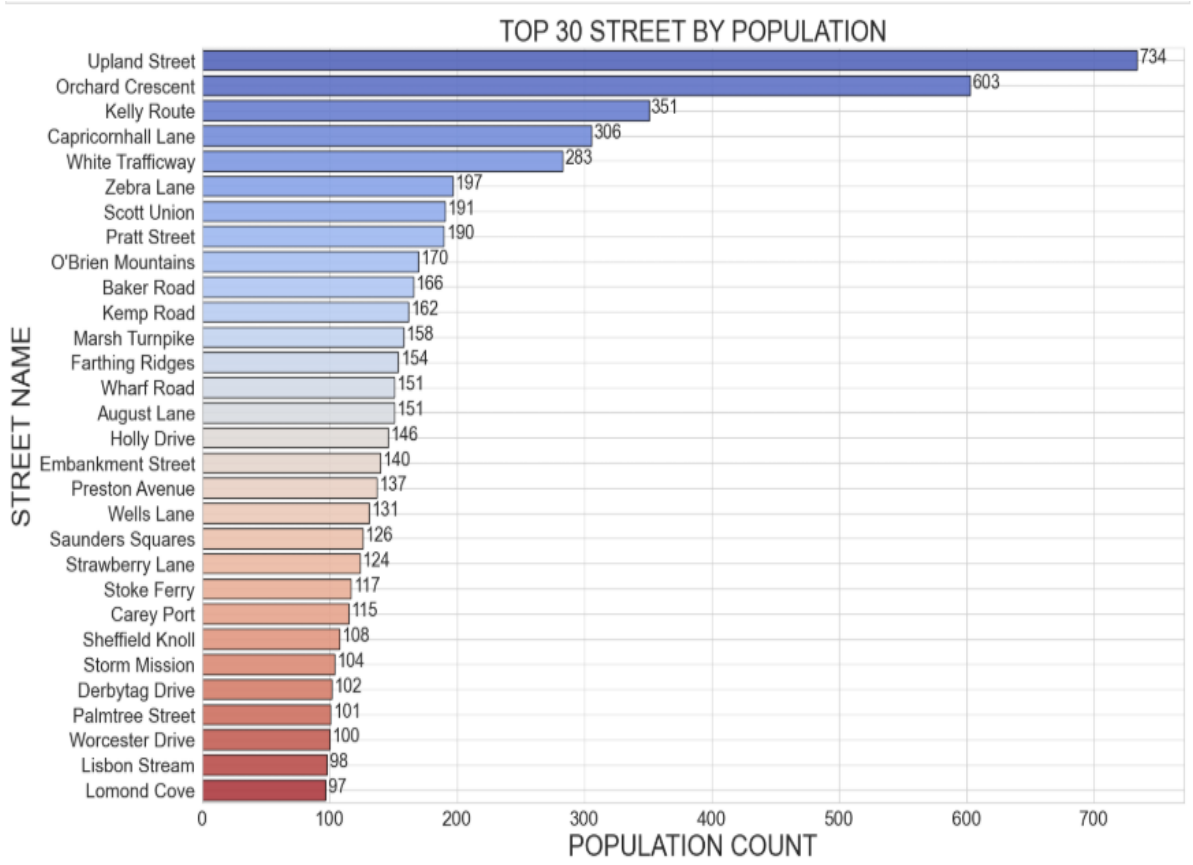


The figure shows the top 30 streets with the most houses. On average, each street has

about 27 houses, with some exceeding this figure. The highest recorded household occupancy is 22 individuals, indicating possible shared or multi-family living.

High house concentrations may increase population density and place greater demand on local infrastructure and services.

Population Density



The above figure shows the distribution of residents across the town's top 30 most populated streets. The top five—Upland Street, Orchard Crescent, Kelly Route, Capricornhall Lane, and White Trafficway—account for approximately 26.8% of the town's total population. The most densely populated street, Upland Street, contains 734 individuals, while the average population per street is around 96.

Although the top five streets do not house the majority of the town's population, they do represent significant residential clusters. This concentration may warrant secondary investment to reduce pressure on existing services. Strategic decongestion through low-density housing or infrastructure expansion around these streets could help balance population distribution more evenly across the town.

Birth and Death Rates : A Comparative Analysis

Understanding birth and death rates helps forecast future service needs.The birth rate, calculated using individuals aged 0–1 (VonVille, no date), is 21.10 per 1000 people. Five years earlier, it was 14.35, showing a 0.60% increase—a modest but steady population growth.The death rate was estimated using an average life expectancy of 75 years, resulting in 4.55 deaths per 1000 people. This low rate suggests a relatively young and healthy community.

The general fertility rate, measuring births per 1000 women aged 15–44, is 82.30—a moderate figure typical of communities balancing work, education, and family life.

Recommendation and Suggestions

Based on the detailed analysis of the town's census data, the following recommendations are proposed to support future development, workforce readiness, and sustainable urban planning:

Investment Priorities

The town’s unemployment rate, calculated at approximately 8.6%, is significantly higher than the national UK average of 3.7% according to Watson (2023). This suggests a need for urgent investment in employment and training programs. Additionally, the town has a large student population and a high proportion of young adults who are likely to enter the workforce soon, further straining job markets and available social services.

According to the National Health Survey for England (Devitt et al., no date), young adults are classified as those aged 16–24 years. Given that students and young adults represent a substantial portion of the town’s population, initiatives focused on skills development, re-training, and youth empowerment are essential.

What Should Be Built on the Unoccupied Plot?

In addressing the council’s request for a proposal, the recommended project for the vacant land is the construction of a Community College.

Building a Community College would offer several strategic benefits:

- It would provide affordable higher education options for the growing population of college-aged students (16–18 years), who currently form a significant percentage of the town’s demographic structure.
- It would reduce the number of university students commuting to neighboring towns or cities, minimizing strain on regional transport systems.

- It would create local jobs, both in education and auxiliary services, boosting employment.
- It would offer skill acquisition and retraining programs for unemployed adults, aligning education with the town's evolving labor market needs.

Research from *The Modern Community College: The Underused Asset in Socioeconomic Development* (2021) highlights that local community colleges play a critical role in creating jobs and boosting economic growth. In many small to mid-sized towns, community colleges are among the largest employers, making them key drivers of local development.

Thus, building a Community College would directly support workforce development, reduce unemployment, and enhance the town's overall economic vitality.

Reference List

Devitt, R., *et al.* (no date) *National Health Survey for England – Young Adults Definition*. NHS England. Available at: https://www.nhs.uk (Accessed: 29 April 2025).

Office for National Statistics (ONS) (2023) *UK Labour Market Overview: March 2023*. Available at: https://www.ons.gov.uk (Accessed: 29 April 2025).

Office for National Statistics (ONS) (2021) *Religion, England and Wales: Census 2021*. Available at: https://www.ons.gov.uk (Accessed: 29 April 2025).

Pew Research Center (2020) *Teens and Religion*. Available at: https://www.pewresearch.org (Accessed: 29 April 2025).

The Modern Community College (2021) *The Underused Asset in Socioeconomic Development*. Lumina Foundation. Available at: https://www.luminafoundation.org (Accessed: 29 April 2025).

VonVille, H. (no date) *Basic Public Health Statistics: Calculating Birth and Death Rates*. Texas Medical Center Library. Available at: https://library.tmc.edu (Accessed: 29 April 2025).

Watson, C. (2023) *UK Unemployment Trends and Regional Comparison*. UK Labour Market Reports. Available at: https://www.uklabour.gov (Accessed: 29 April 2025).