# CENSUS PROJECT REPORT

## Abstract

A detailed analysis of a moderately sized town existing between two much larger cities. The lack of knowledge about the composition of residents in town has led to the local government's inability to make proper land use plans and allocate certain investments. The census analysis aims to recommend what should be built on the unoccupied plot of land and investment to explore. Data was cleaned efficiently and useful insights were sourced. The town has a slightly declining population with a peak at 40-44 age groups and a low birth rate with a high overall life expectancy across the population. Also, the town consists of about 79% commuters. However, the population's life expectancy will only increase provided there is a better living condition such as effective transportation and care homes.

## Introduction

### Background

A census involves the collection and researching of the demographic, economic, and social information relating to all individuals in a country or a part of a country at a specified time usually every 10 years.

Just like most things, a census has both its advantages and disadvantages(Garikaib, 2016). The Census statistics are the most accurate overview of the country's population which some developmental planning such as land use and house planning, demographics trend, assessing socio-economic conditions can be based upon. Also, it has a larger scope of a country's data and characteristics as well as other variables such as occupation trends, income, housing, ethnic group distribution, and religious trends. All these data and information help policymakers and the government to effectively and efficiently decide resources distribution of social amenities/infrastructures and identify social forms and disparities relating to race, ethics, and religion.

A major disadvantage is that household heads can sometimes provide false information thus compromising the data in the census. This can be tackled through efficient data analysis to detect and remove or replace this false information according to other household data provided in the census.

### Problem statement

The lack of knowledge about the population and distribution of residents in communities has made policymakers inability to make proper land-use plans, develop policies, and allocate certain funding and investments. Also, it has led to some residents suffering from the disparity

of resources distribution in the community. Inequalities associated with race, ethics, and religion in addition to disadvantaged groups such as the disabled and the poor are often seen in these communities.

**Objectives**

The main objectives of the population census analysis are to recommend:

1. What should be built on the unoccupied plot of land that the local government wishes to develop; and
2. what type of foundation should be invested in? (employment and training, care homes, general infrastructure, etc.)

**Research questions**

To address these objectives, the following questions are formulated for this research:

1. What is the age distribution of residents in the census?
2. What is the average family size of residents in the census?
3. Who are the commuters in the census?
4. What are the unemployment trends in the census?
5. What is the religious composition of residents in the census?
6. What is the possible population growth in the future? (birth and death rate)

## Methodology

This is a population census of a modestly sized town located between two larger cities in the UK. In this project, a 5-phase technique was used to perform the census data analysis.

Image: G2

The phrase 1 has been fully discussed in the problem statement.

### Data collection

The data used in this project for the census analysis was sourced from here.

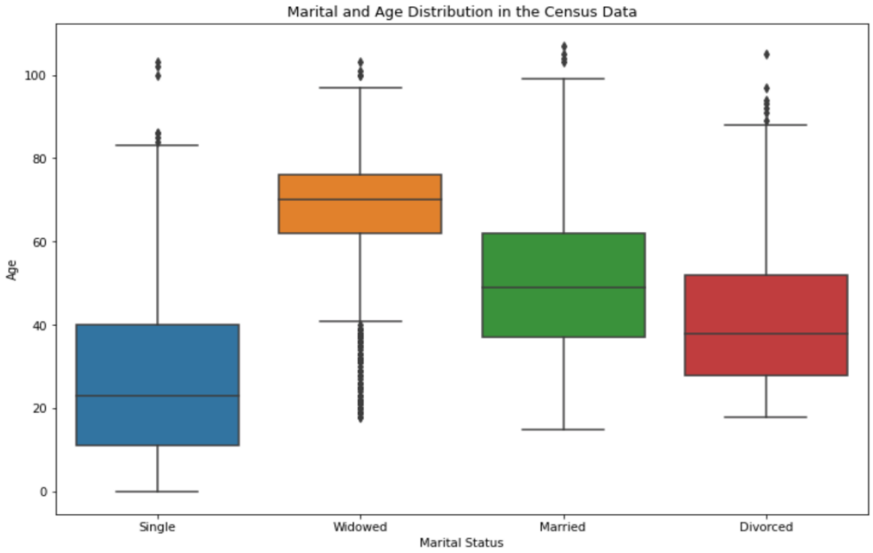
### Data Cleaning

To efficiently clean the data, we have to understand the data. Understanding the data involves various operations such as data loading, data description, and features, data quality, unique values, and outliers, etc. All these operations were carried out effectively with the Pandas library.

The census data consisted of 11 features which are House Number, Street, First Name, Surname, Age, Relationship to Head of House, Marital status, Gender, Occupation, Infirmity, and Religion. Now let's discuss each feature with data errors and how the errors were fixed:

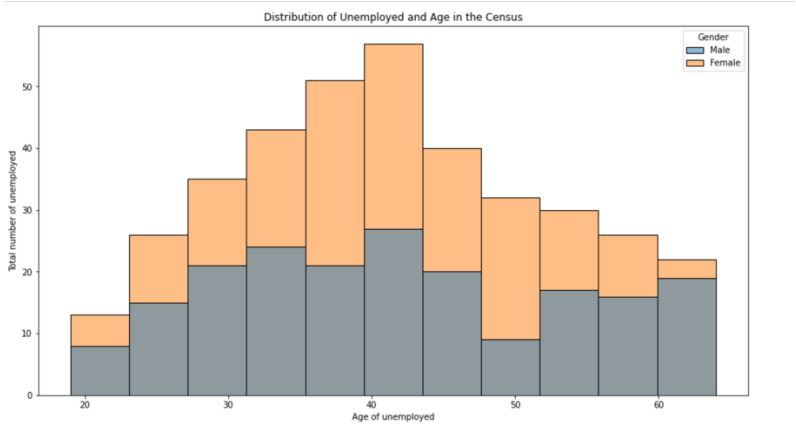
Blank spaces and some missing data were imputed by inferring information from an individual in their household. Features like Surname, House Number, Street, and Relationship to Head of House were fixed with this method. Entries that could not be imputed with referential information were imputed to 'None' (in the case of Religion and Infirmity). 10 entries in Infirmity with blank spaces were replaced with None. Also, one record of infirmity as female was imputed as none. Furthermore, in relationship to the head of house feature, Adopted-son, Adopted-daughter, Adopted-granddaughter were imputed as son, daughter, and granddaughter since they have the same meaning/definition.

Marital status and gender consisted of incorrect data (misspelled words and improper representation of information), these columns were fixed by correcting the spellings. Also, 2247 missing entries found in Marital status were replaced with ‘single’ as the data points fall within the age 0 to 18 (ONS, 2021) which set the minimum marriage age at 16 with the consent of parents or guardians and 18 without that consent.

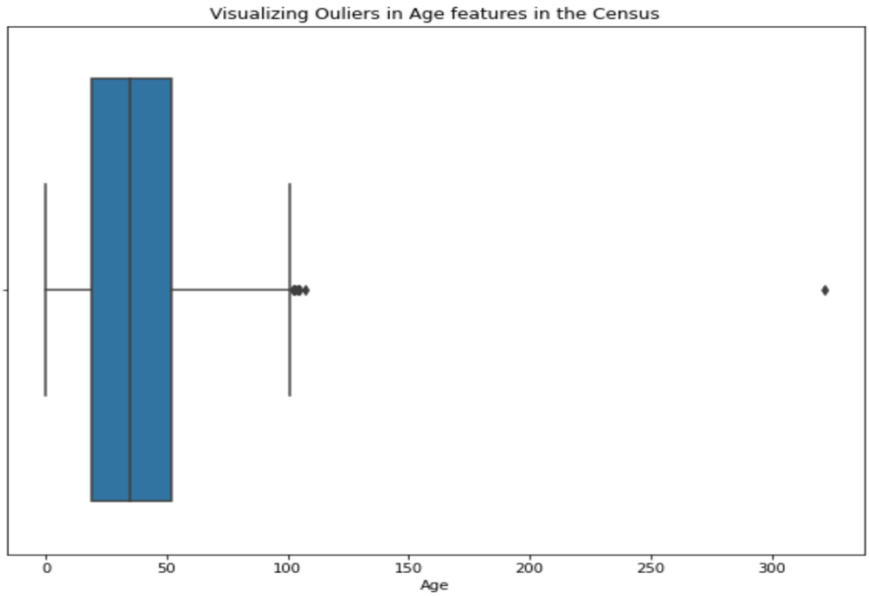



Similarly, several none entries in Religion was falls within the age range 0 to 18. This shows m ost parent in the census does not conform to the ideology of religious transmission from par ent to child. Six entries in religion were replaced with ‘None’ – Female, Quaker, Undecided, J edi, Private, Sith as these are all made-up names and they do not conform to the official relig ions recognized by the UK government(Study UK, 2021). Quaker, Jedi, and Sith are misinform ed information as this was only in 3 entries and can be termed as a lie or joke. 'Females' can be viewed as an individual error when filling the census. Also, undecided and private have th e same meaning as unknown or none which is a perfect fit for the entry replacement.

Two entries with Occupation as blank were imputed as a student, this information was inferred from the age data as both entries are 7 and 8 years old. Although some individuals above 65 years old are unemployed but for this analysis that is valid as forced retirement age of 65 no longer exists according to the UK government(Gov.UK, 2021).



Outliers were removed for one record of age at 322 has no human as lived that long (the age was imputed to be 32 because this entry is a male lodger and has a career in health, I assumed it was an imputation error by repeating the figure 2 twice). Also, a record of age -1 was replaced with 1 as this entry is a twin child, this information was inferred from the household data and was fixed appropriately.




Once the census data has been cleaned, the data features should have the following features:

```
data.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9687 entries, 0 to 9686
Data columns (total 13 columns):
 #   Column                          Non-Null Count  Dtype
---  ------                          --------------  -----
 0   House Number                    9687 non-null   int64
 1   Street                          9687 non-null   object
 2   First Name                      9687 non-null   object
 3   Surname                         9687 non-null   object
 4   Age                             9687 non-null   int64
 5   Relationship to Head of House   9687 non-null   object
 6   Marital Status                  9687 non-null   object
 7   Gender                          9687 non-null   object
 8   Occupation                      9687 non-null   object
 9   Infirmity                       9687 non-null   object
 10  Religion                        9687 non-null   object
 11  Employment Status               9687 non-null   object
 12  Household                       9687 non-null   object
dtypes: int64(2), object(11)
memory usage: 984.0+ KB
```

For effective analysis, the following features were added:

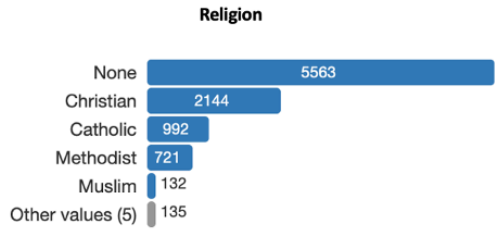
- **Employment status:** Grouping the occupation status into 5 categories: Student (Child), Student, Employed, Unemployed, and Retired.
- **Household distribution:** A combination of street and house number to understand the average family size of residents and possible housing structure.

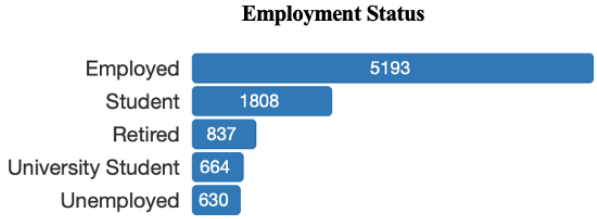
Now that the census data has been cleaned and fixed with the appropriate method, we can go ahead to draw insights and analysis to make a proper recommendation.

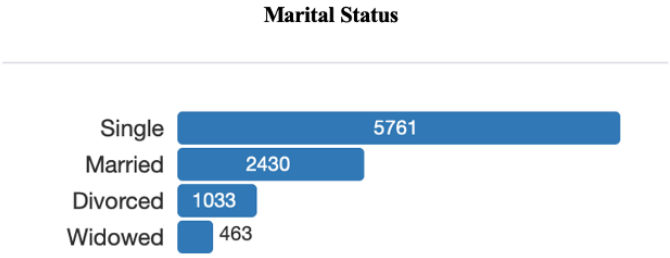
## Results and Discussion

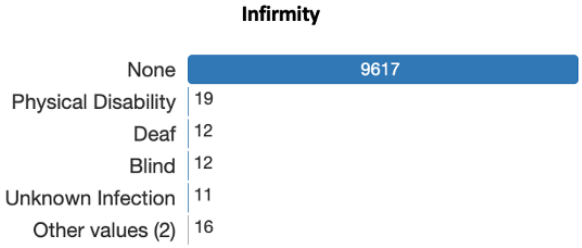
### Detailed analysis

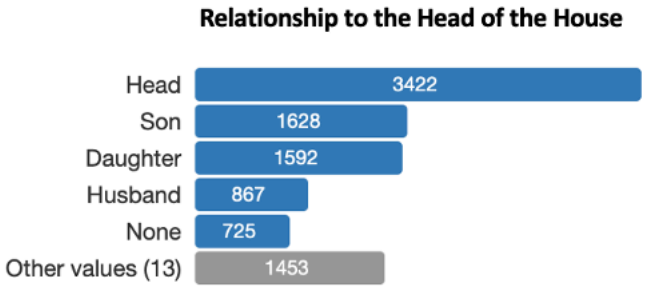
The data collected during the project were analyzed by exploring the descriptive analysis of our data demographics which involves data visualization to view the patterns and relationships amongst our features. The descriptive analysis shows most of the town is employed, with a high number of university students and around 6.5% of the population unemployed. Most of the population are single or married and infirmity in the town is low compared to other features; individuals with an infirmity constituted less than 1% of the overall population:



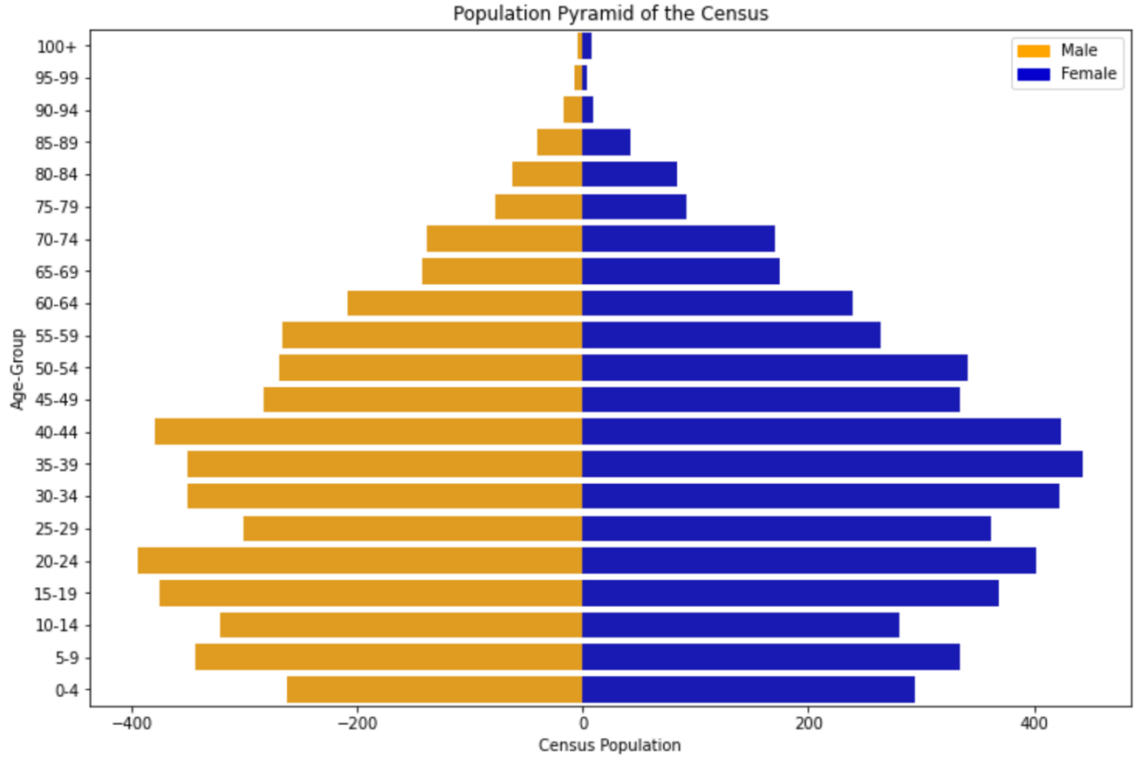






**Age distribution in the census**

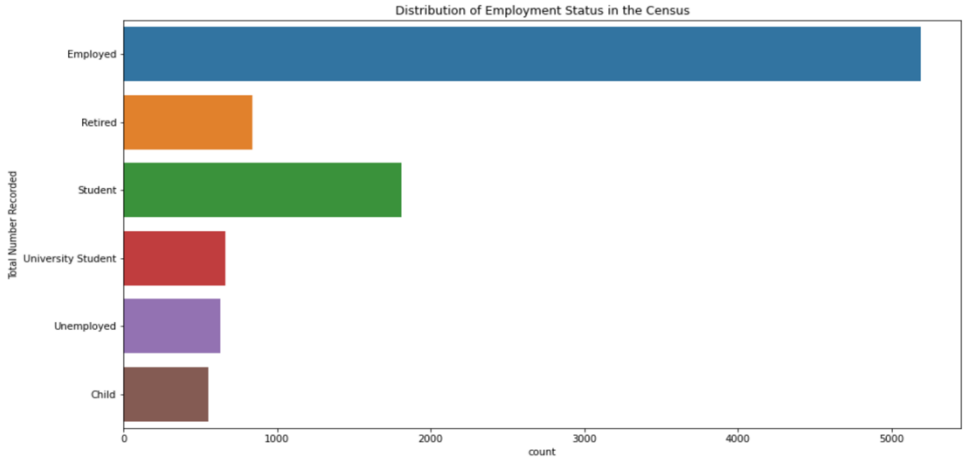
This composition of ages in the census was analyzed by exploring the population pyramid of respondents' age and gender. Below is a chart showing how the ages were distributed in the census.




From the above chart, we observed that the census is a growing population with the highest distribution among the middle-aged group of 40-44 years and the lowest distribution is among the old-aged group of 95-99 years. This distribution shows a lower number of young people compared to middle-aged, mostly individuals aged 0-4. We can propose a low birth rate in the census and a high overall life expectancy for both males and females across the population.

**Commuters in the census**

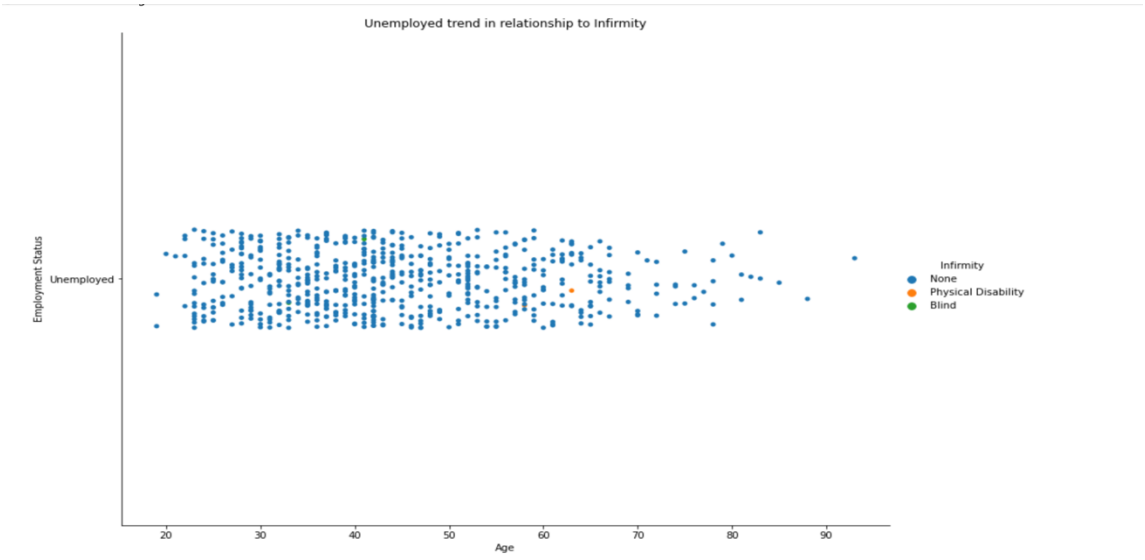
From the pyramid population above, we observed the peak of the pyramid to be among middle-aged, young individuals and lodgers, this shows a working-age group which consists of mainly employed and students. This shows there is a presence of commuters in the census that are commuting to work, school, or both. To get a better overview of the commuter's distribution, let's visualize the occupation trends in the census data.




The chart above shows employed, a university student and student to be the highest occupation in the census accounting for 79% of the total population.

**Unemployment trends in the census**

The above chart on employment distribution reveals that the community has a relatively small record of unemployment compared to other trends. Hence, we need to ask more questions to be sure the reason for unemployment is not a result of health issues. To get an overview, let's visualize unemployment with infirmity.



From the employment status chart above, the number of unemployed with infirmity is very low, thus, unemployed individuals are unskilled or inexperienced persons that should be considered for training.

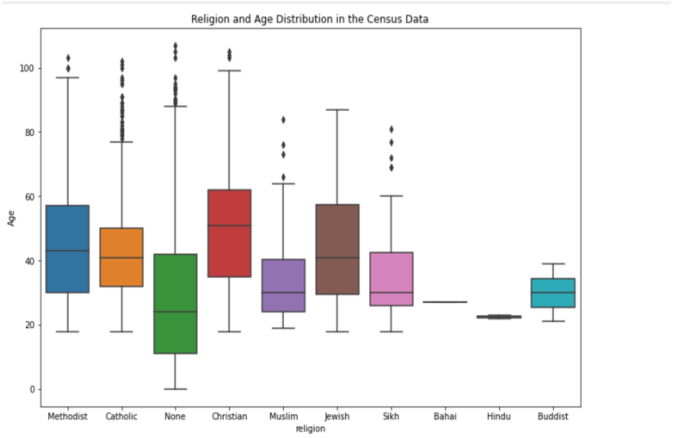
**Average family size of residents in the census**

```
: grouped_household.describe()

: count    3422.000000
  mean        2.830801
  std         2.181412
  min         1.000000
  25%         1.000000
  50%         2.000000
  75%         4.000000
  max        22.000000
  Name: Household, dtype: float64
```

Household distribution was derived by grouping street and house number to understand the average family size of residents per household. The chart above shows the average family size in my census to be 2.83 which is slightly higher than the average family size of residents in the UK which is 2.4 (ONS, 2020). The maximum family size is 22, although this could be as a result of some houses with the same street and house number we can still make a recommendation that better housing structure should be adopted. Considering the slightly declined population, we need to build more low-density houses to aid better living conditions for the residents.

**Religious composition of residents in the census**




**Count and Median Age Per Religion**

| Religion | Age Count | Median Age |
|---|---|---|
| Methodist | 721 | 43 |
| Catholic | 992 | 41 |
| Christian | 2144 | 51 |
| Muslim | 132 | 30 |
| Jewish | 59 | 41 |
| Sikh | 71 | 30 |
| Bahai | 1 | 27 |
| Hindu | 2 | 22 |
| Buddist | 2 | 30 |
| None | 5563 | 24 |

Most religions are still growing in the community as individuals under the age of 18 did not register any religion in the census. Christianity is still the dominant religion (22%), most individuals didn't specify their denomination but we can assume it falls under Methodist, Catholic, and Baptist as these are the denomination recognized by the UK government (Study UK, 2021).

Further analysis on which denomination is dominant will require data specifying this issue. Since other religions comprised a small percentage, though other religions may grow in the future, they do not have a significant following in the town to warrant building a new religious center above other needs of the population.

**Population growth in the future**

**A. Birth rate**

$$\text{Birthrate} = \frac{\text{Births in 1 year}}{\text{Ten thousand total population}}$$

Analysis of the birth rate of the town is observed to be declining. The current birth rate is calculated to be 97 births per ten thousand (annum) while the birth rate in the past 5 years is 115 births per ten thousand (annum). This shows a decline in the rate with a decrease of 18 births. Projecting into the next 5 years the birth rate would have a decrease of 0.18% ten thousand which depicts a declining population growth in the future, although this might not be the case as birth is determined by the human factor (socio) and this can be difficult to analyze or predict.

**B. Death rate**

$$\text{Death rate} = \frac{\text{Deaths in 1 year}}{\text{Ten thousand total population}}$$

The death rate is calculated by estimating deaths by the difference in age bands for those over 65. We then sum up the differences and divide by 5 (to account for the age-banding), the death rate is calculated as 149 deaths per ten thousand. Although factors such as migration can cause this rate we will assume individuals over 65 are most likely to be retired and settled.

C. **Growth rate**

This is calculated by getting the difference between the birth and death rate. The growth rate of the population census is -5.2% per ten thousand, this shows a relatively small growth of the town's population yearly.

**Migration Rate**

Although university students can constituents to the immigration and emigration in the community, we can also deduce immigration statistics from lodgers and visitors who are

single. While emigration statistics are calculated from the difference in male and female divorcees since there is a likeliness of an individual leaving the partner's town after divorce. Considering this method, the immigration rate per ten thousand is 352, and the emigration rate per ten thousand is 1442. This shows we have more people migrating out of that town than people coming into the town.

## Conclusion and Recommendation

As the commuters account for more than half of the population (79.1%), we can recommend an effective transportation (train) to be built on the land to increase the standard of living of the population by allowing easy and smooth commuting to their respective destination Also, since the population slightly decreased, a better housing structure such as building low-density housing should be adopted. Furthermore, the high number of lodgers in the community shows people are migrating to the community and will possibly need rooms to sublet.

Considering the population tends to live well into old age although the old aged distribution is small this can be improved by introducing a better health care system. Thus, investing in aged care should be prioritized. Also, since unemployment is not conditioned on health issue, introducing a training facility will equip these individuals with opportunities to explore skilled jobs and improve their standard of living. Services such as recreational centers should be invested in as the community is comprised of age groups that can benefit from these infrastructures. Apart from providing comfort and better living conditions, these infrastructures will generate more income for the community, encourage individuals and companies to migrate and settle in the community, and this will allow the community to investments in more infrastructures as they grow.

Services such as university and other general infrastructure should not be explored at the moment due to the population's relatively small growth yearly. However, the population's life expectancy will only increase provided there is a better living condition (effective transport and old care homes). I believe these recommendations will improve the overall living conditions of the residents in the town.

## References

Garikaib (2016) 'Advantages and Disadvantages of a Population Census', *Free ZIMSEC & Cambridge Revision Notes*, 28 January. Available at: https://revision.co.zw/advantages-and-disadvantages-of-a-population-census/ (Accessed: 8 December 2021).

Gov.UK (2021) *Check your State Pension age*, *GOV.UK*. Available at: https://www.gov.uk/state-pension-age (Accessed: 9 December 2021).

ONS (2020) *Families and households in the UK - Office for National Statistics*. Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/families/bulletins/familiesandhouseholds/2020 (Accessed: 8 December 2021).

Study, UK. (2021) *Religion | British Council*. Available at: https://study-uk.britishcouncil.org/moving-uk/student-life/religion (Accessed: 8 December 2021).

Wikipedia (2021) 'Age of Marriage Act 1929', *Wikipedia*. Available at: https://en.wikipedia.org/w/index.php?title=Age_of_Marriage_Act_1929&oldid=1048193415 (Accessed: 8 December 2021).