

---

# CENSUS DEMOGRAPHIC REPORT

---

Edem Kwesi Dziko - 202254152



APRIL 28, 2022

---

## Contents

Introduction .....	2
Data Cleaning .....	2
Detailed Analysis & Insights .....	4
Infirmity.....	4
Religion .....	4
Age .....	5
Household Occupancy .....	6
Marital Status.....	6
Occupation, Unemployment & Commuters .....	7
Population Growth.....	8
References .....	10

## Introduction

This report aims to discuss the demographics of a small town in 1881 UK. The town, which is located between two major cities, lacks a university and a train station but does have a Catholic worship centre. There are 8374 people living there, some of whom commute every day to and from the town.

The report seeks to provide details on the data cleaning process through filling up empty spaces, null entries and correcting administrative errors which were recorded during the census. It further provides analysis and insights such as household occupancy rate, unemployment, infirmity, immigration rate etc. These analyses aim at helping the Local Government decide on what to do with a vacant parcel of land and what investment opportunities could be pursued at the end of the study.

Data cleaning and analysis was done solely using the Python Programming Language - through the Jupyter Environment. Some libraries used in processing the census data include Pandas, NumPy, ydata profiling, Seaborn & Matplotlib. To fully understand the Data Cleaning and analysis performed on the Census Data, a copy of the Jupyter Notebook file has been attached for your perusal.

## Data Cleaning

The data was particularly laced with some white spaces, Null entries, and some errors. To correct these, 24 whitespaces were mostly handled by inferring to the entry's household to determine what the missing value is. For example, missing Surnames, Gender and Marital Status was easy to determine from the entry's household, however it was impossible to determine one case of a missing surname which related to an adopted daughter as well as a first name of an 88-year-old man hence, these were imputed as 'Unknown'. The Occupation whitespace was an individual aged above the pension age of 66 (GOV.UK, 2023) therefore, it was imputed as 'Retired Unknown' thus assuming the individual is not self-employed. For the 13 whitespaces in the infirmity column;

```
House Number      0
Street            0
First Name        1
Surname           2
Age               1
Relationship to Head of House  0
Marital Status    1
Gender            3
Occupation        1
Infirmity         13
Religion          2
dtype: int64
```

Fig.1– data whitespaces

```
House Number      0
Street            0
First Name        0
Surname           0
Age               0
Relationship to Head of House  0
Marital Status    1975
Gender            0
Occupation        0
Infirmity         0
Religion          2013
dtype: int64
```

Fig.2– Data Null Values

underestimating the importance of empty infirmity would be risky, consequently they were labelled as 'Unknown' for further investigations to be carried out.

Furthermore, there were significant NaN entries in both the Marital Status and Religion columns. Out of the 1975 NaN entries in Marital Status as shown in Fig.2 only 1 entry was above 17 years hence was converted as 'Single' after a house search. All others (minors) were converted as 'Single-minors' to differentiate between adult singles. Nonetheless, there were 4 children

deemed to be head of house who were also married in the data set. These homes were removed from the dataset because it contained two substantial inconsistencies—a child marriage which is no longer permissible even with parental consent (GOV.UK, 2023) and a child serving as head of the household. These records were not deemed important enough to the broader study to allow for multiple counts of inaccurate data to be imputed to the entire household.

Also, since children in the UK are allowed to be free from religious affiliations (UNCRC, Article 14) 'NaN' Religion entries associated with children were converted to 'Undecided'. This has the effect of making it impossible to analyse how religion is transmitted from parents to children. Alternatively,

adults, whom were either students, PhD students or university students were imputed as 'None' based on the mode religion for these group of people after household inferences proved futile.

To prepare the data for data analysis and insights, the following steps were carried out:

- The presence of FLOAT values mixed with INT values in the Age Column. The FLOAT values were subsequently converted to INT.
- Housekeeper and Private were seen as computational errors and corrected as 'None' in the Religion column after house searches.
- Similarly, it will be counterproductive to include Jedi as a religion because Jediism stems from mythological Star Wars movie. (BBC UK, 2016)
- An 'Age\_Band' column was created as part of the dataset which grouped individuals in 5-year groups.
- 'None' entries in Relationship to Head of House were changed to Tenants.
- The main Occupation column was further classified into 6 sub-groups namely, Employed, Child Student, Uni Student, Unemployed, Children and Retired. This new classification became necessary since, there were various job roles under the Occupation column, again, retired workers still had 'Retired' adjoined to their previous roles. In doing this, all children aged below 5, were named as children, retired workers were named as Retired, University Students and PhD Students were named as Uni Student and children aged between 4 and 18 were kept as Child Student.
- A column was created for household occupancy to aid in further analysis.
- There are 12 religion groupings within the town. Out of these, groups with representation of less than 1% were reclassified as 'Others' thus Sikh, Jewish, Buddhist, Quaker, Pagan & Orthodoxy.

At the end of the data cleaning process, the dataset had the following features and statistical summaries:

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 8364 entries, 0 to 8376
Data columns (total 14 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   House Number                        8364 non-null   int64
1   Street                             8364 non-null   object
2   First Name                         8364 non-null   object
3   Surname                           8364 non-null   object
4   Age                               8364 non-null   int32
5   Relationship to Head of House      8364 non-null   object
6   Marital Status                    8364 non-null   object
7   Gender                            8364 non-null   object
8   Occupation                        8364 non-null   object
9   Infirmary                        8364 non-null   object
10  Religion                          8364 non-null   object
11  Age_Bands                        8364 non-null   category
12  Occup_Classification             8364 non-null   object
13  Num_Occupants                   8364 non-null   int64
dtypes: category(1), int32(1), int64(2), object(10)
memory usage: 1.1+ MB
```

Fig.3– Data features

Note: 'Num\_Occupants' should not be used for any statistical purposes as it would provide misleading

	count	mean	std	min	25%	50%	75%	max
House Number	8364.0	31.192970	34.259426	1.0	9.0	20.0	40.0	197.0
Age	8364.0	36.335127	21.780068	0.0	19.0	36.0	51.0	116.0
Num_Occupants	8364.0	4.096604	3.031304	1.0	2.0	4.0	5.0	22.0

information when used in calculations. However, it aided in plotting graphs related to number of household occupants. 'num\_occupants' on the other

Fig.4– Data statistics

hand can be used for statistical purposes related to household occupancy.

## Detailed Analysis & Insights

### Infirmity

The healthy population (labelled as 'None') was dropped in plotting the Infirmity graph in order not to distort the graph.

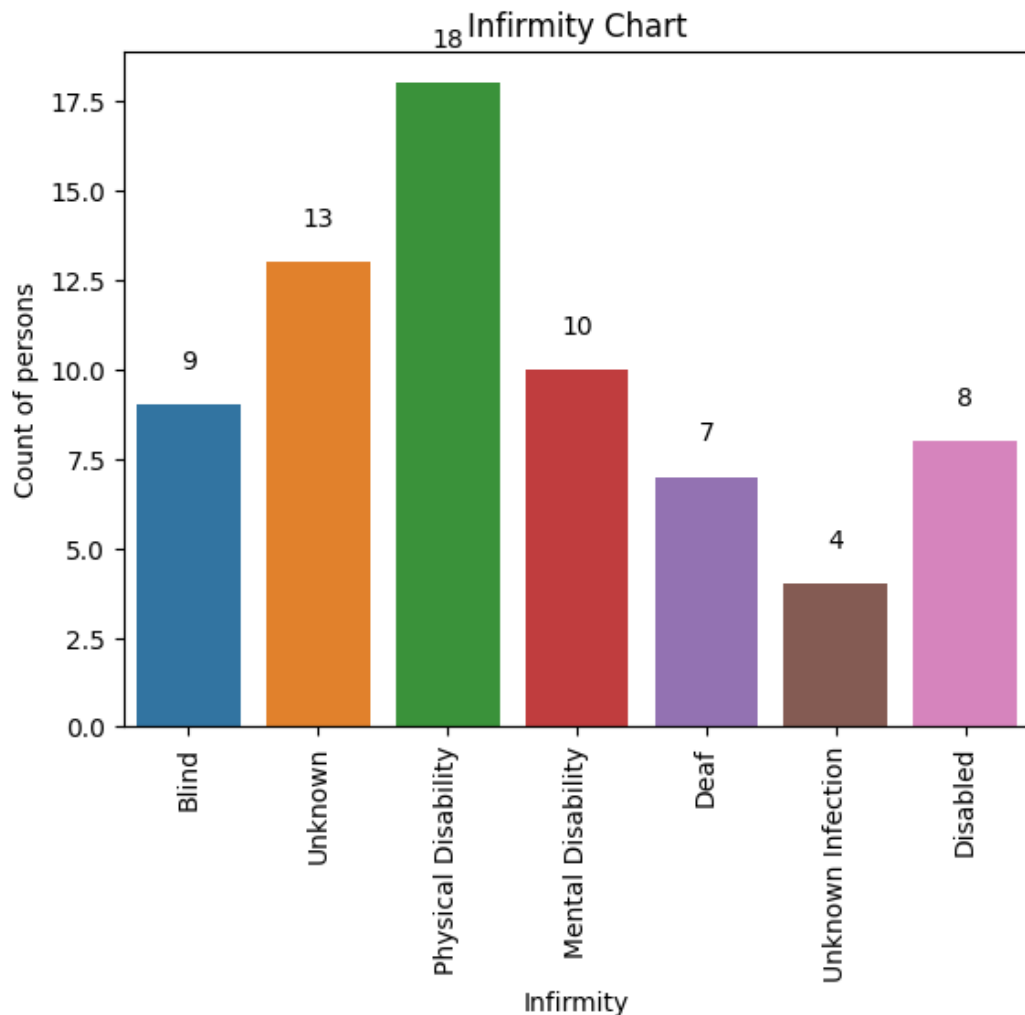


Fig.5– Infirmity Chart

From a distance, the populace appears to be in good health, with only 0.82% of the population with disabilities. However, it is crucial to provide further medical care to the four patients with unknown infections to stop contagious transmission, if any. Similarly, the 13 people tagged as unknown need to be contacted to ascertain their true status.

### Religion

Although having no designated place of worship, the Christian community is the most prevalent faith in the town. There are also a lot of nonreligious persons as well as young people who aren't sure which faith to follow. Fig.6 below depicts the smaller religious groups; thus, Muslim group and the 'Others' category as having increasing population. Also, the presence of many outliers among Catholics indicates a significant number of people over the age of 80.

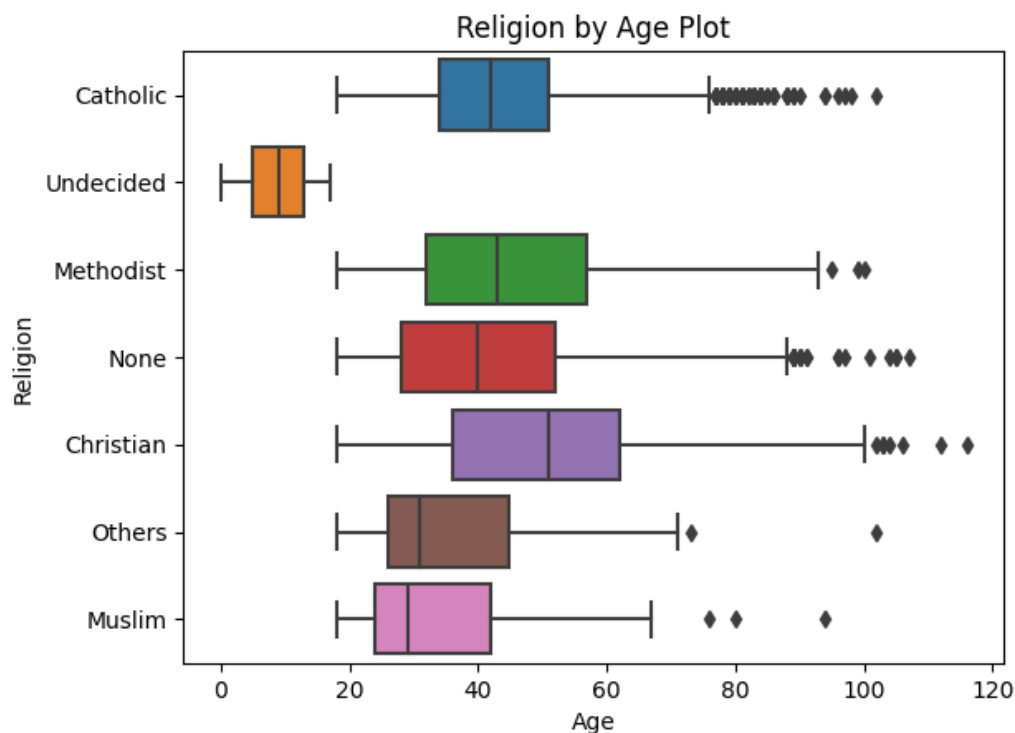


Fig.6– Religion by Age plot

## Age

The age analysis below demonstrates that the school-age and 35-50 age groups are more densely populated. This demonstrates without further analysis, the need for a train station to help ease movements between the town and other neighbouring places. However, it is premature to decide.

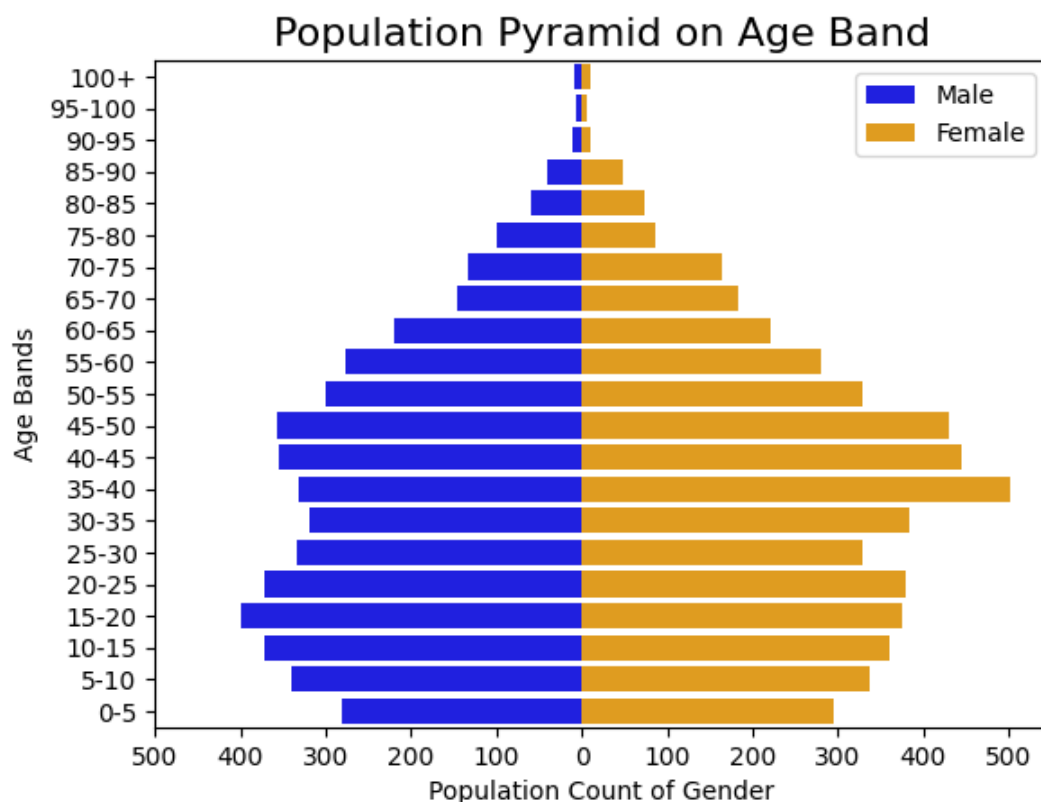


Fig.7– Age Pyramid

It is also evident that the number of children continues to decrease throughout all age groups, from 15-20, to 10-15, 5-10 & 0-5. This illustrates the downward nature of the town's birth rate. Likewise, population has been on the descend for the aging residents. This may be directly related to mortality, but more research is needed to determine whether other factors, such as migration or divorce, also contribute.

## Household Occupancy

The housing occupancy analysis is rather an interesting one. There are a total of 3060 homes within the town. Although the average occupants per household (2.7) 3 is not significantly different from the

count	3060.000000
mean	2.733333
std	1.930672
min	1.000000
25%	1.000000
50%	2.000000
75%	4.000000
max	22.000000
Name:	num_occupants,

Fig.8– Household Statistics

UK average occupancy rate of (2.4) 2 (ONS, 2022). A closer look at the data reveals some worrying findings. As shown in Fig.9 below; the two oldest people in the town live in a 21-occupancy house which is detrimental to their well-being. Secondly, about 200 people aged over 80 years reside in one or two occupancy homes. These older people, most of whom are married, single or widowed, should not be left alone because of their age. Overall, 29.4% of all residences (901) in the town are typically occupied by one individual. This had a significant impact on the calculation of the average occupancy rate. Additionally, the proportion of households with occupancy levels equal to or below the norm is 72.5%, meaning that 27.5% of the town's households, or 4274 people making up 51.1% of the total population, are overcrowded. Also, 237 homeowners have listed their residences for rent. There are 599 tenants living in these houses with an average occupancy of 2.53.

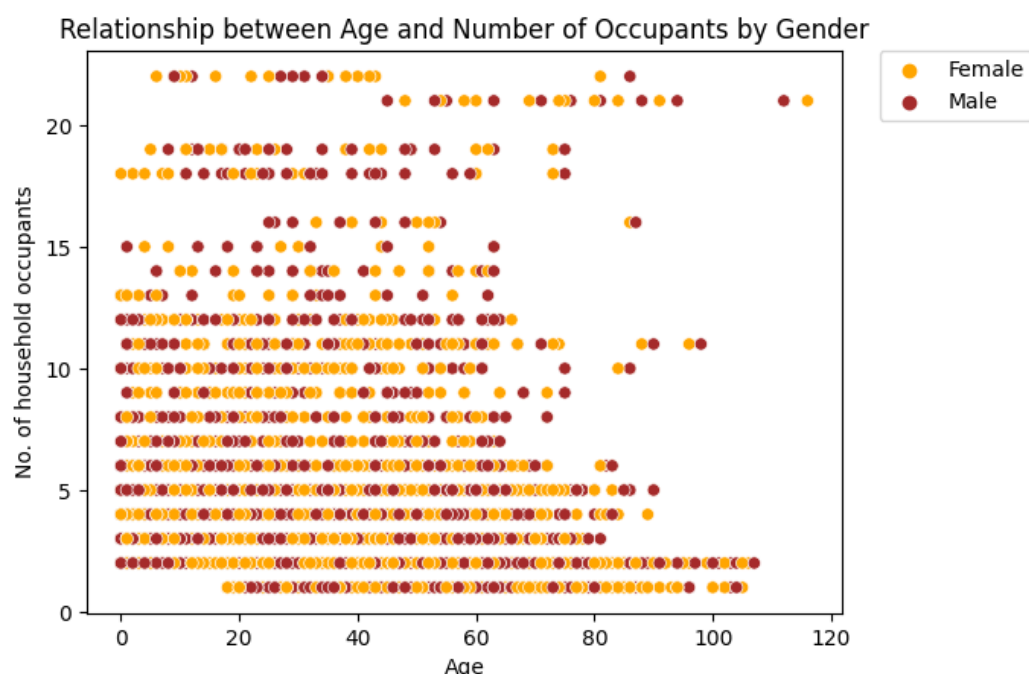


Fig.9– Household Statistics

## Marital Status

Apart from Single-minors, the existence of outliers can be noticed for each item in the Fig.10 below. In contrast, the widowed plot is intriguing since it exposes more outliers prior to the first quartile than subsequent quartiles. This indicates the presence of widows under the age of 48 and even under the age of 20. The Single storyline, which depicts outliers who are single after the age of 80, is another fascinating tale. The town's marriage rate is 347 marriages per 1000.

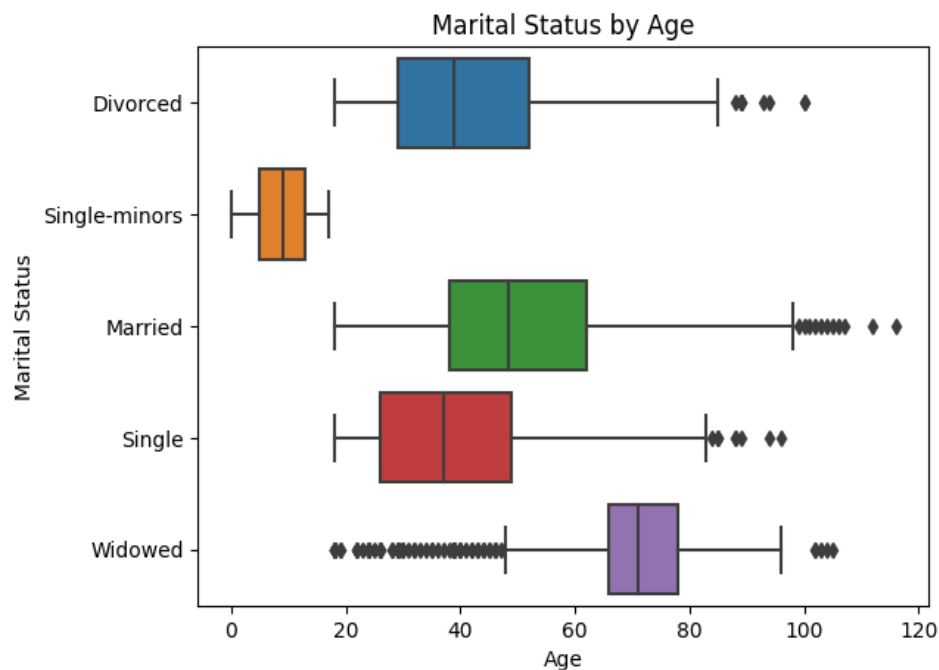


Fig.10 –Marital Status by Age

In terms of divorcees, there are 473 female divorcees against 317 male divorcees, leaving 156 unaccounted-for divorcees. The divorcee emigration rate was calculated by finding the difference between the female and male divorcees by the total population. Emigration stood at 19 emigrants per 1000. Note: this assumption is slightly flawed because differences could be as a result of death and not necessarily emigration.

	Count	Percentage
Gender		
Female	473.0	60.0%
Male	317.0	40.0%
Total	790.0	100.0%

Fig.11– Divorcee Stats

## Occupation, Unemployment & Commuters

Employed persons within the town make up 54% of the town's population. Whereas 8.5% of the working age and 5.8% of the overall population are unemployed, respectively. These rates continue to be higher than the 3.8% current unemployment rate in the UK. (ONS, 2023)

Job roles including but not limited to the following: accommodation manager, academic librarian, administrators, aid worker, amenity horticulturist, bookseller, buyer retail, community pharmacist, secretary and education officer could all be jobs within the town hence does not require commuting to nearby cities. Assuming that, 128 distinct occupations thus 20% of unique job roles within the employed category, are located within the town, this equates to 899 (20% of 4495) non-commuting workers. Overall, the commuter population was projected to be 49.45% (3596 commuting workers and 540 Uni Students) of the total population. consequently, this situation demonstrates the intense pressure that exists on the road system connecting the other cities.

	Occup Classification	Count	Percentage
0	Employed	4495	53.7
1	Child Student	1581	18.9
2	Retired	791	9.5
3	Uni Student	540	6.5
4	Unemployed	487	5.8
5	Children	470	5.6

Fig.12– Occupation Stats



## Population Growth

To arrive at the population growth rate, the following must be examined:

- a. **Birth Rate:** Since there is no prior census data to base your comparison on, the birth rate was computed using the crude birth rate and the age-specific birth rate. According to the crude technique, the current birth rate is 10 births per 1000, compared to 13 births per 1000 in the census taken five years earlier. From this, the overall birth rate has decreased by 3 births per 1000. Using the age specific birth rate which provides a more accurate figure; birth rate was calculated using 20-25 and 25-30 age groups. This resulted in a birth rate of 226 and 319 births respectively. Birth rate fell by 93 births per 1000 using this method. This birth rate analysis corroborates the decline seen in Fig.7 above.
- b. **Death Rate:** Death Rate was calculated in a similar manner as Birth Rate. Even though the life expectancy for the population in 1881 was acquired to be around 42-45 years for both male and female (ONS, 2022), checks indicate that 36% of the population are aged over 45 years therefore it would be inappropriate to compute death rate from this. Relying on the current life expectancy in the UK (ONS, 2022) the 80-85 and 85-90 age group for analysis, the difference between these groups were divided by the total number of people in the group by a 1000 of population. This was subsequently divided by 5 to arrive at the death rate per year of 39 deaths per 1000. Similarly, in calculating the total death rate from age 60 to 120 years, death rate was computed as 45 deaths per 1000 for these age groups. Note: it is incorrect to presume that there were the same number of people in each of these two generations when they were both the same age. Also, the sheer difference between two groups does not automatically indicate mortality.

	Age-Group	Count_1	Count_2	Pop Size	Number of Death	Death Rate per Annum
0	60-65 & 65-70	439	330	769	109	28.0
1	70-75 & 75-80	299	187	486	112	46.0
2	80-85 & 85-90	133	90	223	43	39.0
3	90-95 & 95-100	22	13	35	9	51.0
4	100-105 & 105-110	13	4	17	9	106.0
5	110-115 & 115-120	1	1	2	0	0.0

Fig.13- 60-120years Deathrate

- c. **Immigration:** The Immigration rate was calculated by identifying the number of people who came into the town by the total populace. Visitors and Lodgers were identified to be the group of people who migrated into town. Immigration rate was 36 per 1000.
- d. **Emigration:** This was calculated based on the difference between female and male divorcees by the total population. Emigration rate for the town is 19 emigrants per 1000 whilst net migration is 17 migrants per 1000.

In summary, the aggregate of emigration and death rate was subtracted from the aggregate of birth rate and immigration. The population growth rate is the result in this case. The population is projected to increase annually by 204 people.

## Recommendations

Considering that commuters make up about half of the entire population, it would be necessary for the Town to have its own train station where commuters would easily travel to work and school. However, this solution does not seem to benefit the entire population and should rather be budgeted for the long-term future.

Examining the fact that, a significant proportion of homes have occupancy rates above the average implies the presence of overcrowding or housing shortages which would have adverse effects on the health and wellbeing of residents. Therefore, it would be rather prudent to construct low density homes on the vacate piece of land which will result in future lower occupancy rates. Furthermore, because the town's essential needs are being satisfied by the increased number of homes and probable low cost of living, additional immigrants will be drawn to it.

Additionally, given that constructing a train station comes second to the aforementioned priority and the potential of future population growth, the local government must make investments to upgrade the town's general infrastructure. As part of a long-term plan, infrastructure can help communities experience economic growth and improved quality of life. (NIC, 2021). Infrastructure investment should particularly focus on improving the current road network in and out of the town, pedestrian, and cycle ways, expanding the water, waste collection and energy systems within the town.

In the future, the Local Government should focus on investing in elderly care which might reduce the emigration of both the elderly and divorcees.

## References

GOV UK (2023) *Legal age of marriage in England and Wales rises to 18*

Available online:

<https://www.gov.uk/government/news/legal-age-of-marriage-in-england-and-wales-rises-to-18#:~:text=The%20Marriage%20and%20Civil%20Partnership,if%20they%20have%20parental%20consent> [Accessed: 18-03-2023]

ONS UK (2022) *Household and resident characteristics, England and Wales: Census 2021*

Available online:

<https://www.ons.gov.uk/peoplepopulationandcommunity/householdcharacteristics/homeinternetandsocialmediausage/bulletins/householdandresidentcharacteristicsenglandandwales/census2021> [Accessed: 30-03-2023]

ONS UK (2023) *Unemployment*

Available online:

<https://www.ons.gov.uk/employmentandlabourmarket/peoplenotinwork/unemployment> [Accessed: 22-03-2023]

GOV UK (2023) *State Pension age Review 2023*

Available online:

<https://www.gov.uk/government/publications/state-pension-age-review-2023-government-report/state-pension-age-review-2023> [Accessed: 8-04-2023]

UNCRC Article 14 *Convention on the Rights of the Child*

Available online:

<https://www.ohchr.org/en/instruments-mechanisms/instruments/convention-rights-child> [Accessed: 8-04-2023]

BBC UK (2016) *Jedi is not a religion, Charity Commission rules*

Available online: <https://www.bbc.co.uk/news/uk-38368526> [Accessed: 8-04-2023]

ONS UK (2022) *Mortality in England and Wales: past and projected trends in average lifespan*

Available online:

<https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/lifeexpectancies/articles/mortalityinenglandandwales/pastandprojectedtrendsinaveragelifespan#:~:text=Both%20period%20life%20expectancy%20at,to%2082.6%20years%20for%20females> [Accessed: 7-04-2023]

NIC UK (2021) *Infrastructure, Towns and Regeneration*

Available online:

<https://nic.org.uk/studies-reports/infrastructure-towns-and-regeneration/infrastructure-towns-regeneration-final-report/#tab-summary> [Accessed: 10-04-2023]