

# Census Project Report

This report evaluates and analyses the census of a modest town to make recommendations and future developments on an unoccupied plot of land as well as town investments. To do this, the census data has been cleaned to correct data errors and missing data.

The census data has then been analysed, exploring the fields, looking for interesting trends to enable visually representation and to support the recommendations provided.

To maintain the original census data, a copy of the original censuses was made to manipulate as needed.

## Data Cleaning

The census data was cleaned to ensure that all errors were corrected and to enable proper evaluation and analysis of the data. All the cleaning can be found in the corresponding Jupyter Notebook for reference.

Before delving into the data cleaning, the dataset was checked to identify the data types of the feature. The data types are object and float. The 'Age' column's data type is float, this was corrected and changed to integer for correct analytic use.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10171 entries, 0 to 10170
Data columns (total 11 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   House Number                          10171 non-null  object
1   Street                                10171 non-null  object
2   First Name                             10171 non-null  object
3   Surname                                10171 non-null  object
4   Age                                    10171 non-null  float64
5   Relationship to Head of House          10171 non-null  object
6   Marital Status                         7735 non-null   object
7   Gender                                 10171 non-null  object
8   Occupation                             10171 non-null  object
9   Infirmary                             10171 non-null  object
10  Religion                               7673 non-null   object
dtypes: float64(1), object(10)
memory usage: 874.2+ KB
```

At initial view, two columns can be seen to have missing values: Marital Status and Religion. There are also blank data which were sourced out by checking the features individually.

```
: House Number      0
  Street            0
  First Name        0
  Surname           0
  Age               0
  Relationship to Head of House  0
  Marital Status    2436
  Gender            0
  Occupation        0
  Infirmary         0
  Religion          2498
dtype: int64
```

Missing and Blank data were imputed by referring and gathering information from the individual's data and household information – in the case of House Number, Street, Surname, Relationship to Head of House, and Gender. Any data that could not be imputed through this method, were imputed as 'None' (in the case of 'Infirmary').

There were two records in which the 'First Name' was missing. To correct this, the individuals' surname was imputed as their 'First Name'.

The Religion column had five empty cells. There were also some spelling mistakes which were corrected: Cath0lic – corrected to Catholic, Christ – corrected to Christianity, Christian changed to Christianity. Two religions were converted to Christianity: 'Baptist' and 'Quaker'. Both religion are Christian religious groups.

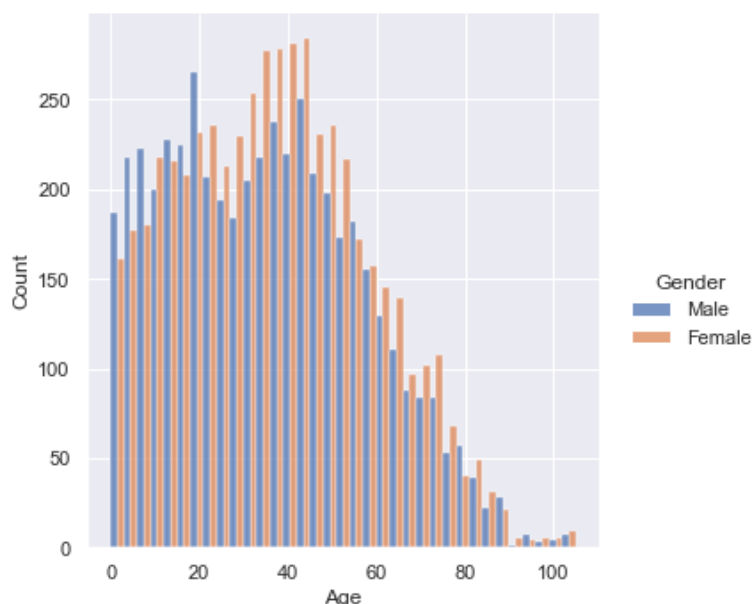
All "NaN" imputes for religions were changed to "None" to make it clearer, except in cases where the individual's religion was changed mimic the household's religious believe. All 'Undecided' and 'Private' entries were also converted to 'None'.

All Religion entries were categorised into the following groups: Christianity', 'None', 'Islam', 'Catholic', 'Methodist', 'Jewish', 'Sikhism', 'Buddhism', 'Bahá'í Faith', 'Christian', 'Orthodoxy'.

Individuals aged under 18 have single as their Marital Status.<sup>1</sup> There is, however, one record of a 16 year old who is married to a 19 year old and have a child. According to the Marriage Act. Section 3, it is possible to get married in the UK if you are under 18, with parental or guardian consent.<sup>2</sup>

Marital Status for individuals aged 18 and under, which were imputed as 'Widowed' were changed to Single – after checking their individual information and the households to gain more information about the young individuals, these records were determined to be incorrect and changed to 'Single' to match the rest of their record.

One incorrect entry in Gender was changed from 'Fem' to 'Female'. There are more Female than there are Male in almost every age.



---

<sup>1</sup> Census 2021. What is your marital or civil partnership status? - Census 2021. [online] Available at: <https://census.gov.uk/ni/help/help-with-the-questions/paper-questions-help/what-is-your-marital-or-same-sex-civil-partnerships-status> [Accessed 4 January 2022].

<sup>2</sup> Legislation.gov.uk. n.d. *Marriage Act 1949, Section 3*. [online] Available at: <https://www.legislation.gov.uk/ukpga/Geo6/12-13-14/76/section/3> [Accessed 4 January 2022].

The data type of 'House Number' was an object/string. This has been converted to integer. There was a missing entry under 'Relationship to Head of House'. After checking the household's information, 'Head' was imputed for the individual as the other person living in the house had imputed 'Partner' as their Relationship to the Head of House.

An incorrect entry in 'Occupation was changed from 'Studen' to 'Student'.

To facilitate the analysis of the dataset, the following columns were added:

**Age Group:** Ages were categorised into 5 year age groups. This was done for the population pyramid.

**Occupation Category:** Occupations were placed into the following groups: Child, Student (including University and PhD Students), Employed, Unemployed, and Retired.

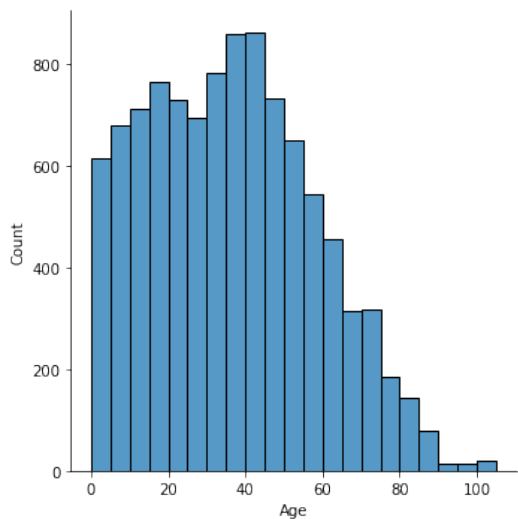
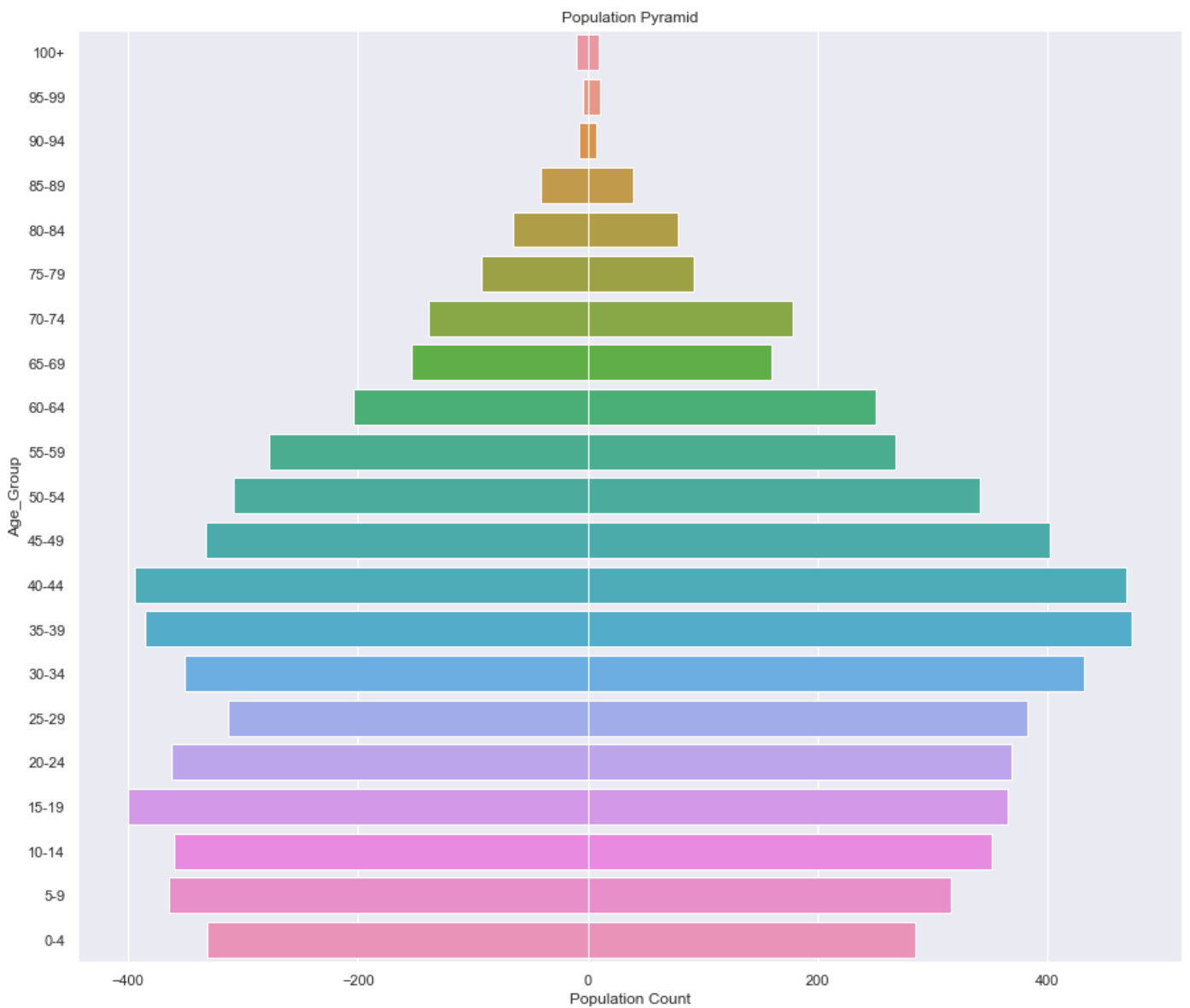
**Household Occupancy:** This represents a count of all individuals in a household. This will be useful when determining over or under occupancy.

As the data has been cleaned, the census data now have the following features and data types:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10171 entries, 0 to 10170
Data columns (total 14 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   House Number                          10171 non-null  int64
1   Street                                10171 non-null  object
2   First Name                            10171 non-null  object
3   Surname                               10171 non-null  object
4   Age                                    10171 non-null  int64
5   Age_Group                             10171 non-null  object
6   Relationship to Head of House          10171 non-null  object
7   Marital Status                        10171 non-null  object
8   Gender                                10171 non-null  object
9   Occupation                            10171 non-null  object
10  Occupation Category                    10171 non-null  object
11  Infirmary                              10171 non-null  object
12  Religion                               10171 non-null  object
13  Household Occupancy                    10171 non-null  int64
dtypes: int64(3), object(11)
memory usage: 1.1+ MB
```

# Data Analysis

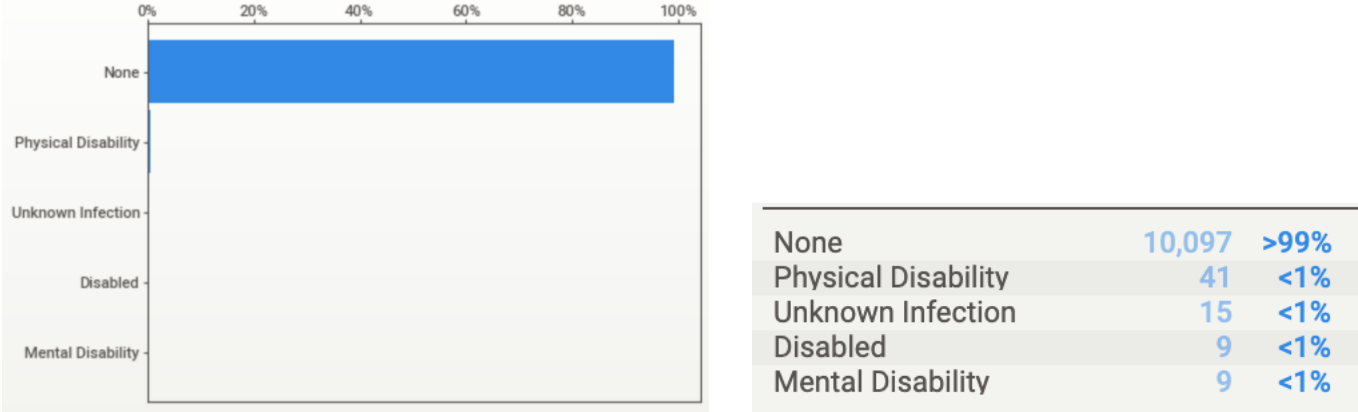
Looking at the population pyramid, it is noticed that there is a lower number of young inhabitants in comparison to the middle-aged. The low number of the population aged 0-4 may suggest a low birth rate in the town.



There is a very high number of middle-aged people – 40-45 in the population. It can also be noticed that the population lives well into old age. The oldest individuals are aged 105, this might be a good reason for the town to invest in or build a care residence facility for the older population.

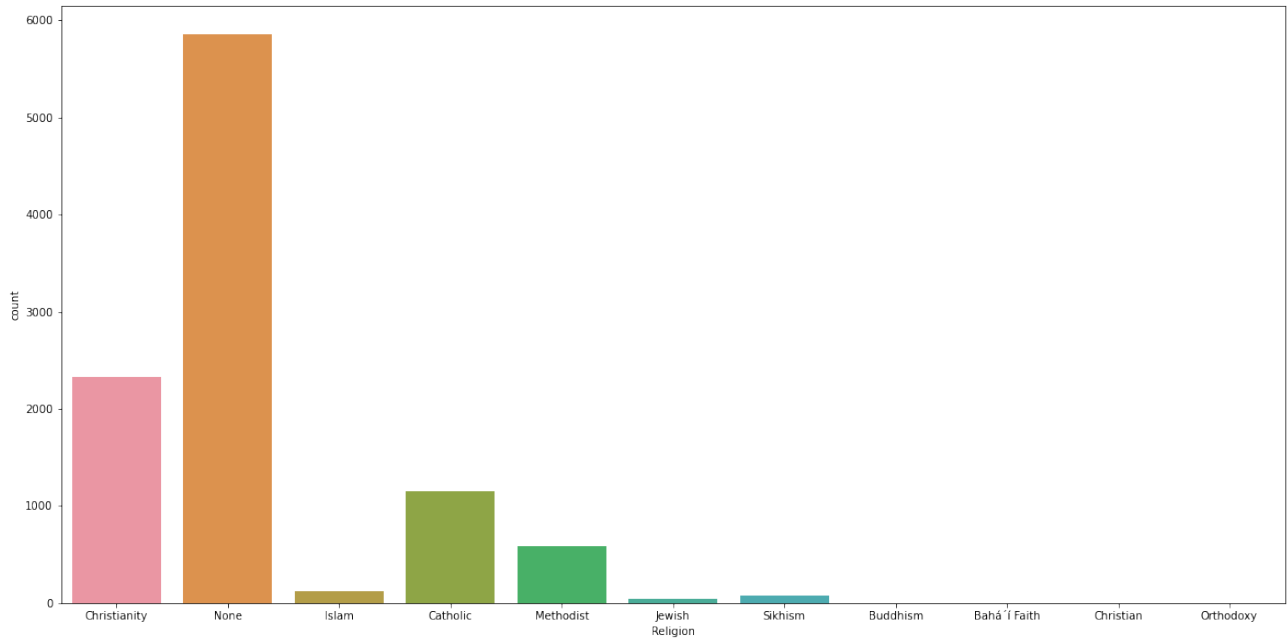
### Infirmity

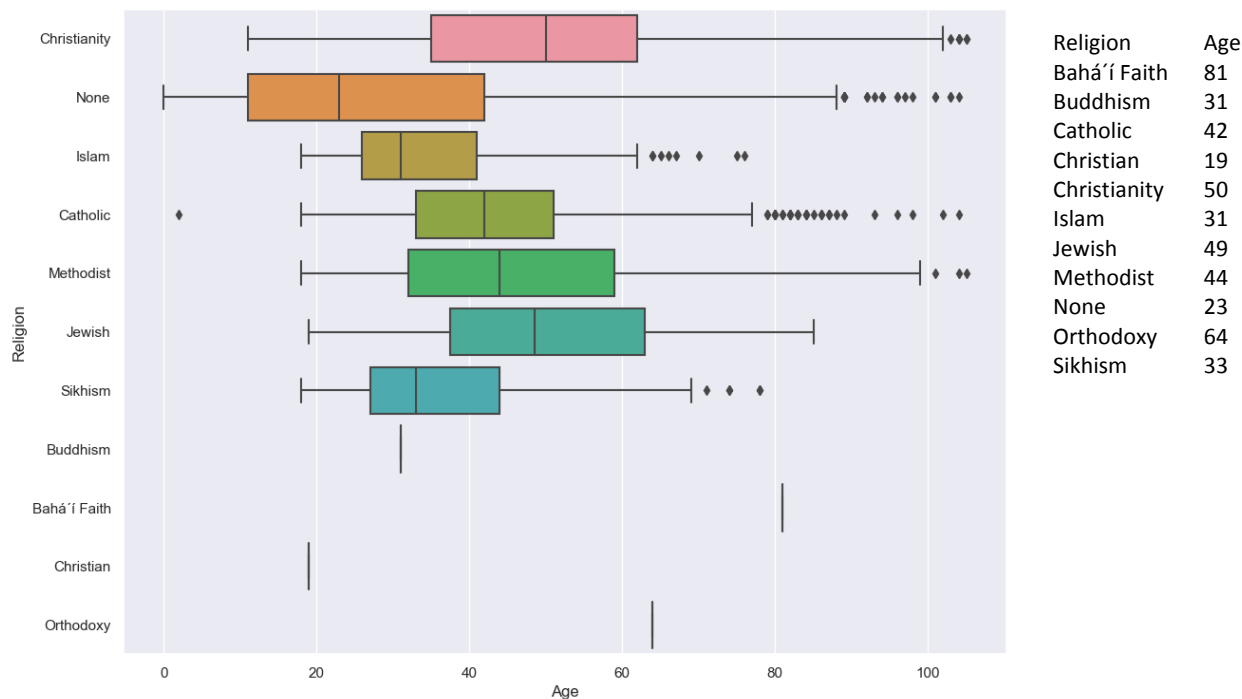
Majority of the population has no infirmity; therefore, no analysis was carried out on this, and no recommendations were made based on the population’s infirmity.



### Religion

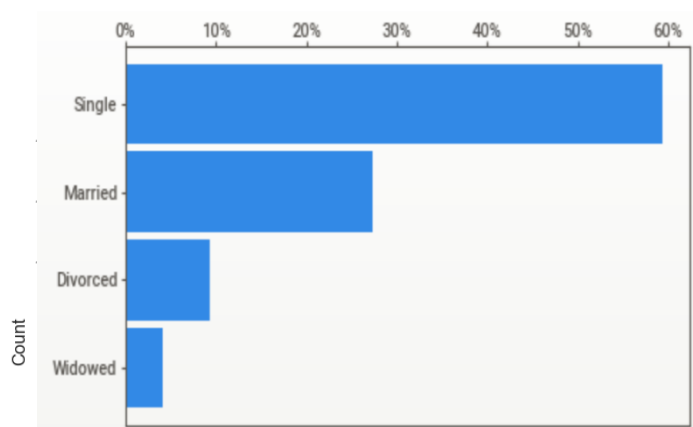
Majority of the population do not have a religion; aside from this, the most popular religion in the town is Christianity (not including Catholic or Methodist), followed by Catholic and Methodist, all of which derive from Christian faith. Catholic is still a popular religion; however, Christianity (other denominations) is the dominant religion in the town, therefore, may warrant building another church for their needs.



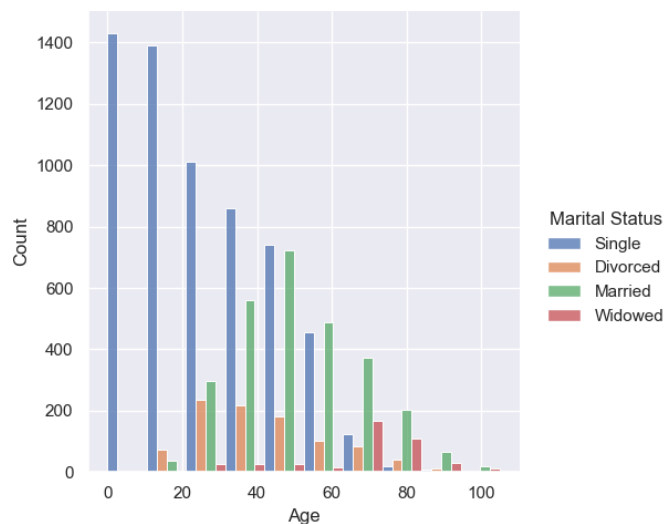
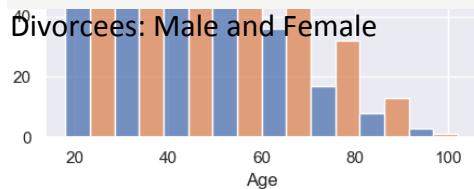


## Marital Status

Majority of the population are single (accounts also for people under 18). Also, marriage and divorce occur from a young age. There are more female divorcees than there are male, this could mean that male divorcees leave town or more female divorcees relocate to the town.



Divorcees: Male and Female



## Birth and Death Rate

Examining the birth and death rate of the town indicates that the town is not growing as compared to previous years.

Calculating the crude birth rate, shows that that there are currently 10 births per thousand, whereas 4 years prior the birth rate was calculated to be 14 births per thousand. The birth rate has fallen by 4 children per thousand in a 4 year period. With this knowledge is can be estimated that

The death rate has been calculated by assessing the difference in age groups for those age over 66 (the current retirement age is 66<sup>3</sup>). People aged 66 and over are most likely retired, without need to migrate, which means the difference in the age groups are more likely to be due to deaths.

The decline amongst other age group is most likely due to migration, students and divorcees moving away from town.

The death rate of the population is calculated to be 11.5 per thousand.

## Migration

There are over 2573 Students (Including University). This means that students are regular in the town and contribute towards the town's growth, even after the University students have graduated and leave town, the younger age group and migration from out of town will occupy the homes left by previous students, take up employments in the town and make use of the town's services, thus consistently attributing to the town's growth.

Immigration is calculated by lodgers who are single and visitors. Divorced lodgers are excluded as they may remain in town lodging after their divorce and therefore should not be classified as immigrating into town. Immigration to the town is calculated as 35 per thousand.

Emigration is calculated from the difference in male and female divorcees, and it estimated to be 19 per thousand.

---

<sup>3</sup> GOV.UK. 2022. *Proposed new timetable for State Pension age increases*. [online] Available at: <<https://www.gov.uk/government/news/proposed-new-timetable-for-state-pension-age-increases>> [Accessed 4 January 2022].

## Commuters

University Students (including PhD students) and majority of those in employment are classified as commuters. Except Students, the unemployed, retired and children, it is estimated that 63% of the population are expected to commute.

1,951	19%	Student
665	7%	Unemployed
616	6%	Child
606	6%	University Student
19	<1%	Toxicologist
16	<1%	Development worker, community
16	<1%	Dramatherapist
16	<1%	PhD Student
16	<1%	Museum/gallery conservator
16	<1%	Engineer, electronics
16	<1%	Chartered loss adjuster
15	<1%	Building surveyor
15	<1%	Dispensing optician
15	<1%	Building control surveyor
15	<1%	Product manager
15	<1%	Race relations officer
15	<1%	Nurse, learning disability
15	<1%	Arts development officer
15	<1%	Health and safety inspector
15	<1%	Health physicist
15	<1%	Surveyor, building
15	<1%	Hydrographic surveyor
15	<1%	Food technologist
15	<1%	Industrial/product designer
14	<1%	Intelligence analyst
14	<1%	Herbalist
14	<1%	Health service manager
14	<1%	Clinical cytogeneticist
14	<1%	Lighting technician, broadcasting/film/video
14	<1%	Interpreter
14	<1%	Designer, jewellery
14	<1%	Charity fundraiser
14	<1%	Chemist, analytical
14	<1%	Sound technician, broadcasting/film/video
14	<1%	Health and safety adviser
14	<1%	Commercial art gallery manager
14	<1%	Data processing manager
14	<1%	Petroleum engineer
14	<1%	Buyer, retail
13	<1%	Financial risk analyst
13	<1%	Haematologist
13	<1%	Equities trader
13	<1%	Nutritional therapist
13	<1%	Psychologist, sport and exercise
13	<1%	Engineer, site
13	<1%	Transport planner
13	<1%	Public affairs consultant
13	<1%	Adult nurse
13	<1%	Sales promotion account executive
13	<1%	Media planner
13	<1%	Structural engineer
13	<1%	Scientist, marine
13	<1%	Comptroller
13	<1%	Surveyor, rural practice
13	<1%	Teaching laboratory technician
13	<1%	Brewing technologist
13	<1%	Engineer, technical sales
13	<1%	Chief Strategy Officer
13	<1%	Metallurgist
13	<1%	Horticultural therapist
5,540	54%	(Other)

Employed	5,469	54%
Student	2,573	25%
Retired	848	8%
Unemployed	665	7%
Child	616	6%

7% of the population are unemployed. This may be due to lack of jobs in the town and jobs being out of town. A train station built in the town may facilitate transportation to jobs out of town and thus lower the unemployment rate of the town.

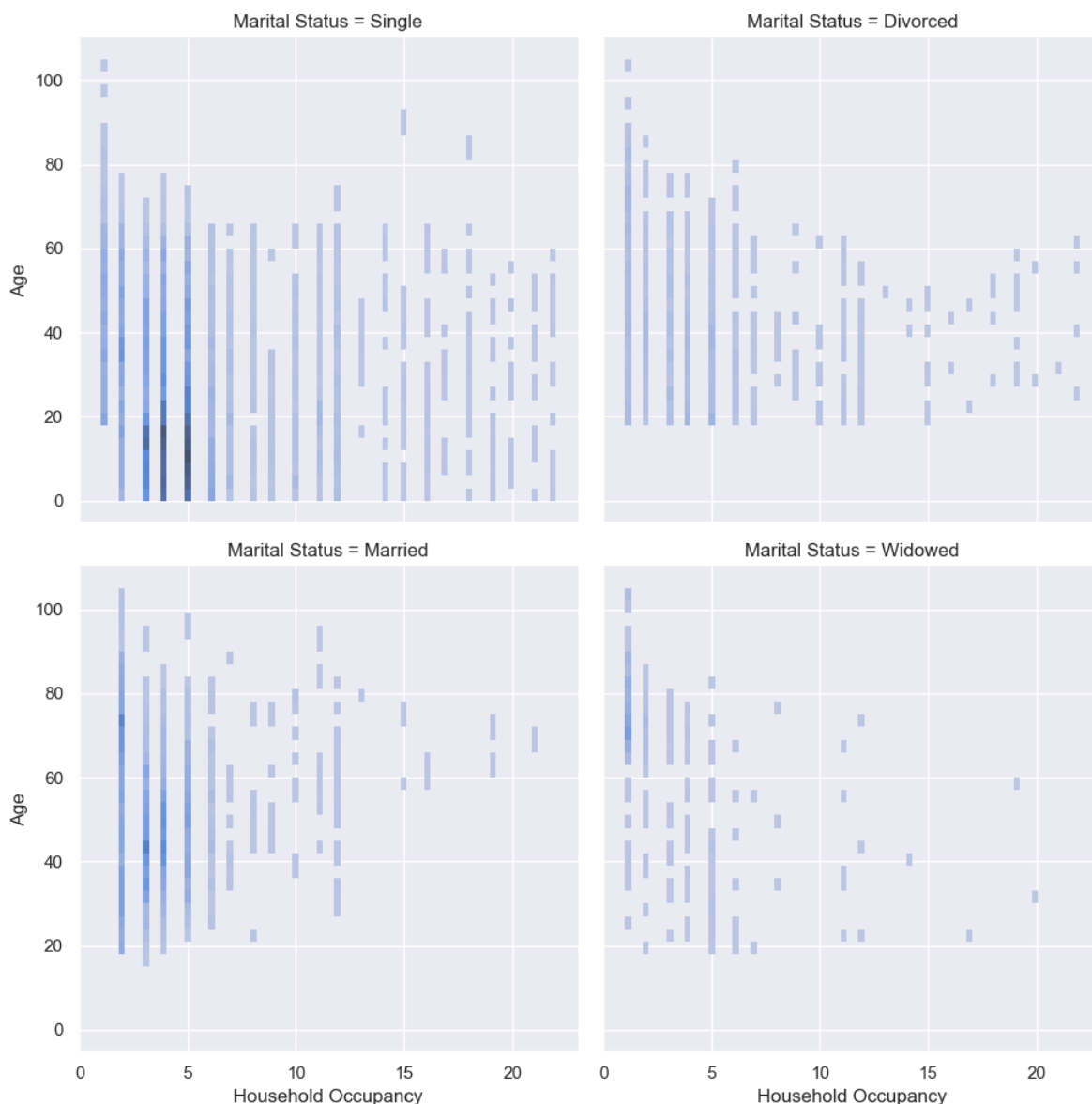


## Household Occupancy Rates

The median household occupancy is four. By this, it can be determined if there is over-occupancy. By this median, there are 3578 over occupied homes. Assuming in a 'big' family household there are 6 people (2 parents and 4 children), it is calculated that there are 1052 with more than 6 occupants and 622 with over 10 occupants. The maximum household occupancy is 22.

These high occupancy could be attributed to families living with students and/or lodgers and visitors.

The plot below shows occupancy by marital status. Married individuals occupy median sized homes of 4 occupants, while single individuals occupy a range of household sizes, with household of 5 being the mostly occupied by under 18 year olds. Divorced individuals also occupy a range of household sizes. Older widowed individuals tend to occupy household of 2 occupants.



Over-occupancy could be due to houses being used to house students – Student accommodation – lodgers and visitors – rented out as hostels, hotels or a bed and breakfast.

## Recommendations

Considering the high number of students and other commuters as well as visitors and lodgers migrating to the town, there is a need for a train station as well as low density housing to be built. Given the high number of over-occupied homes, it should be considered more urgent and beneficial for the families in town and individuals moving to the town, the building of low density housing, instead of the train station, on the unoccupied plot of land. However, in the future, a train station should be developed in the town. This will facilitate the commute for students, the employed – and even the unemployed looking for jobs – and other members of the town.

Given the growing popularity of the Christian religion, a new place of worship – a non-Catholic Church, can be considered in the future, however, it doesn't look to be of significant necessity at the moment. As evidenced above, the members of the population live well into their old age, this suggests that members of the population will live long, and as the aging population increases, they will require a care facility and care services. Therefore, investing in old age care services should take precedence over other services.

## Bibliography

Census 2021. What is your marital or civil partnership status? - Census 2021. [online] Available at: <https://census.gov.uk/ni/help/help-with-the-questions/paper-questions-help/what-is-your-marital-or-same-sex-civil-partnerships-status> [Accessed 4 January 2022].

Legislation.gov.uk. n.d. *Marriage Act 1949, Section 3*. [online] Available at: <https://www.legislation.gov.uk/ukpga/Geo6/12-13-14/76/section/3> [Accessed 4 January 2022].

GOV.UK. 2022. *Proposed new timetable for State Pension age increases*. [online] Available at: <https://www.gov.uk/government/news/proposed-new-timetable-for-state-pension-age-increases> [Accessed 4 January 2022].