Census Project Report

This project report examines census data from a small town sandwiched between two much larger cities connected by highways to make decisions about what to do with an unoccupied plot of land and where to invest. As described in the first section of this report, the census data was cleaned first to correct data errors and missing records before making these decisions.

The sections of the report that follow will highlight key analyses that were conducted specifically to support the decisions made. Following an overview of the town's demographics, an exploratory data analysis and visualization are provided.

Data Cleaning

The Census data was cleaned to correct data errors, and the Jupyter Notebook contains a detailed description of the cleaning process. Empty/missing data cells were imputed and assumed by extrapolating information from an individual's record or from the records of others in their household (in the case of missing first name, age, and gender). Jedi as a Religion was converted to 'None' as it was discovered that Jedi is not a religion. It is a fictional belief system featured in the Star Wars franchise. Religion for minors (under 18) was entered as 'None'. In addition, occupants under the age of 18 have the marital status 'NA'. An exception was made for those over the age of 16, as the legal marriage age in the UK is 16, but those under the age of 18 require parental consent. The dataset was reduced by one household. In this household, a 15-year-old girl was married with an adopted child, listed as the Head of the Household, and had a 19-year-old husband who was a University Student. It looks like an illegal marriage, so the household was dropped from the dataset as it does not have a significant effect on the overall analysis. Age outliers were removed for a couple with the husband being aged 124. It was assumed that they were not sure of their real age due to old age.

Furthermore, three occupants had an empty gender, but further investigation into the origins of their names, which were discovered to be feminine, aided in correctly classifying them as females.

Population Demographics

Following data cleaning, the finalized census data will have the following characteristics:

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 8655 entries, 0 to 8659
Data columns (total 14 columns):
 #   Column                          Non-Null Count  Dtype
---  ------                          --------------  -----
 0   House Number                    8655 non-null   int64
 1   Street                          8655 non-null   object
 2   First Name                      8655 non-null   object
 3   Surname                         8655 non-null   object
 4   Age                             8655 non-null   int64
 5   Relationship to Head of House   8655 non-null   object
 6   Marital Status                  8655 non-null   object
 7   Gender                          8655 non-null   object
 8   Occupation                      8655 non-null   object
 9   Infirmity                       8655 non-null   object
 10  Religion                        8655 non-null   object
 11  age_range                       8655 non-null   object
 12  Employed                        8655 non-null   object
 13  New_Occupation                  8655 non-null   object
dtypes: int64(2), object(12)
memory usage: 1.2+ MB
```

To help with the analysis, the following has been added:

• age_range: Ages placed into 5-year age bands for the population pyramid.

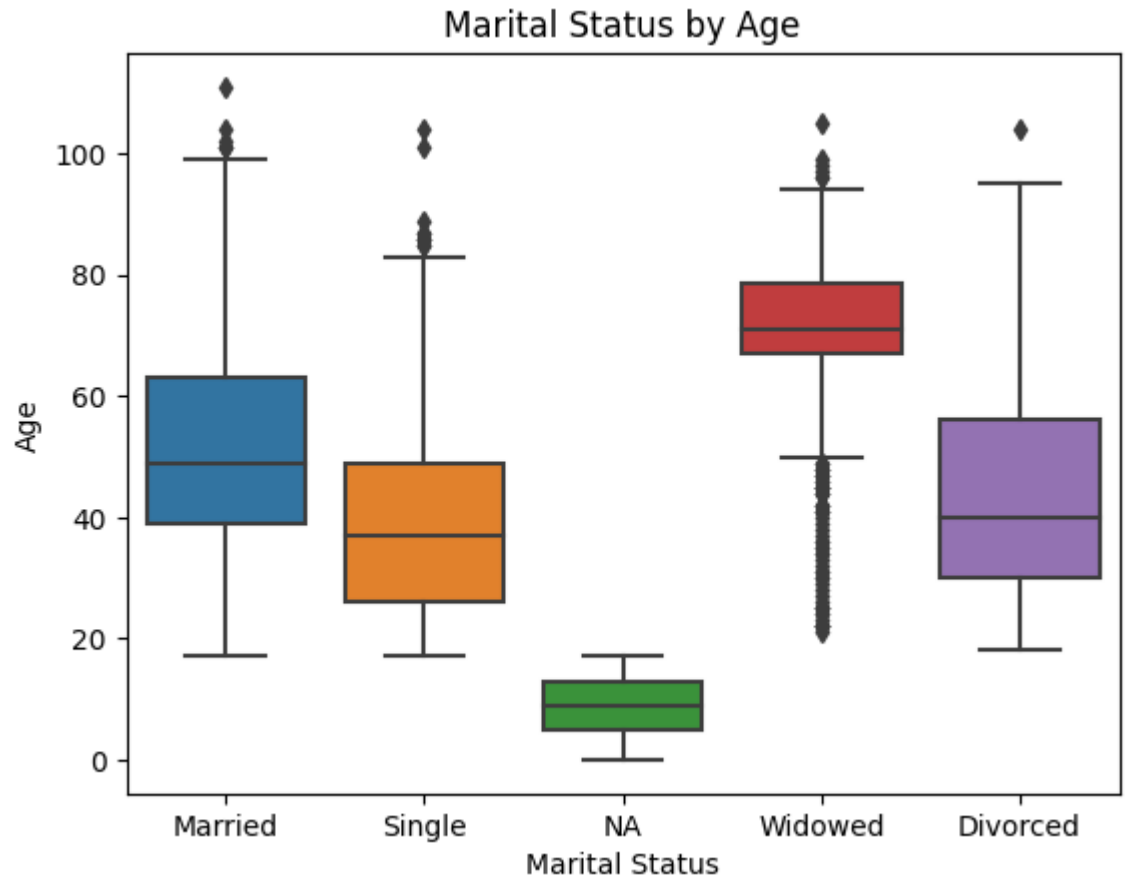
• New_Occupation: Categorized occupations with the values: Employed, Unemployed, Retired, Student, University Student, Child, Ph.D. Student

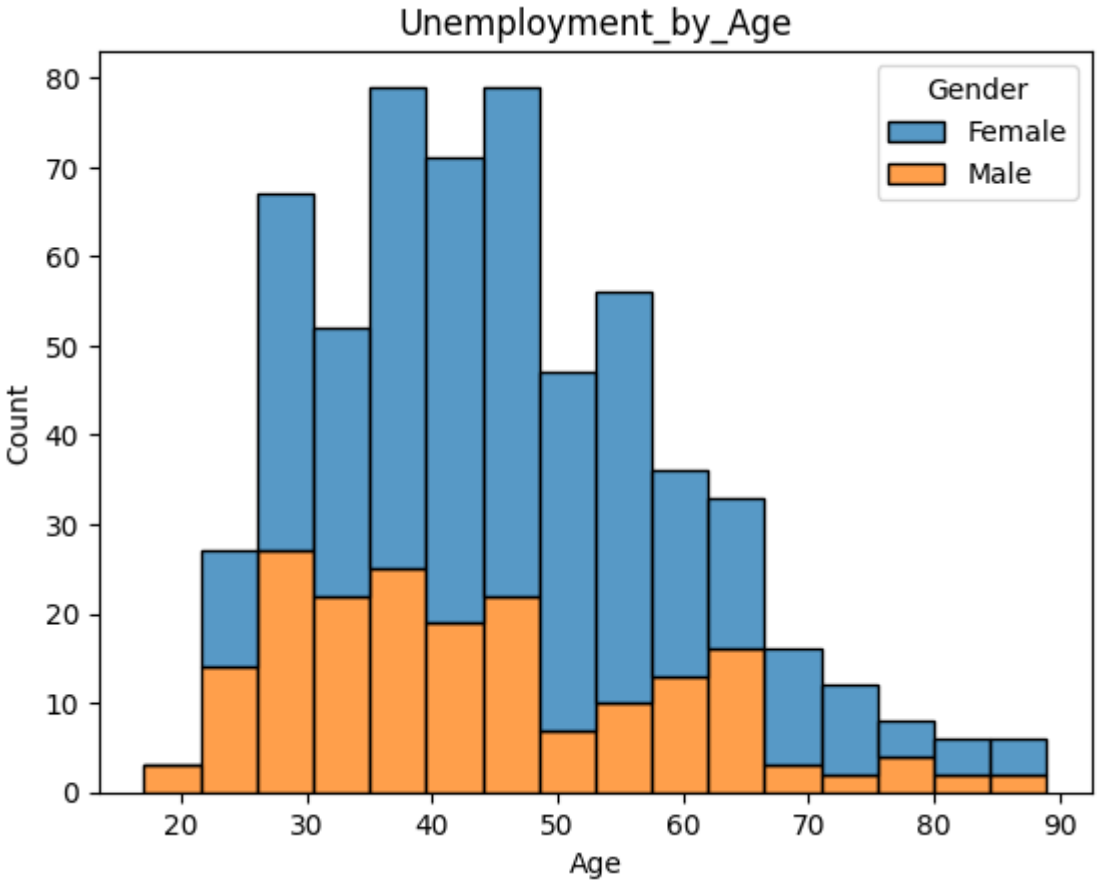
• Household Occupancy (for exploratory data analysis and visualization only): a count of all individuals in a household using their unique street number and house number.

Exploratory Data Analysis and Visualization

Below is a detailed breakdown of Marital Status by age, as well as a boxplot distribution highlighting outliers, most notably 'Widowed' with a standard deviation of 15 and a modal age of 69. Marital Status was also considered 'NA' for those under the age of 16.

	Count	mean	std	min	25%	50%	75%	max	Mode
Marital Status									
Married	2346	51.064365	16.585673	17	39	49	63	111	45
Single	3079	38.109451	14.336893	17	26	37	49	104	18
NA	1879	8.998936	4.975874	0	5	9	13	17	10
Widowed	471	69.210191	15.026515	21	67	71	78.5	105	69
Divorced	880	43.557955	17.922061	18	30	40	56	104	30

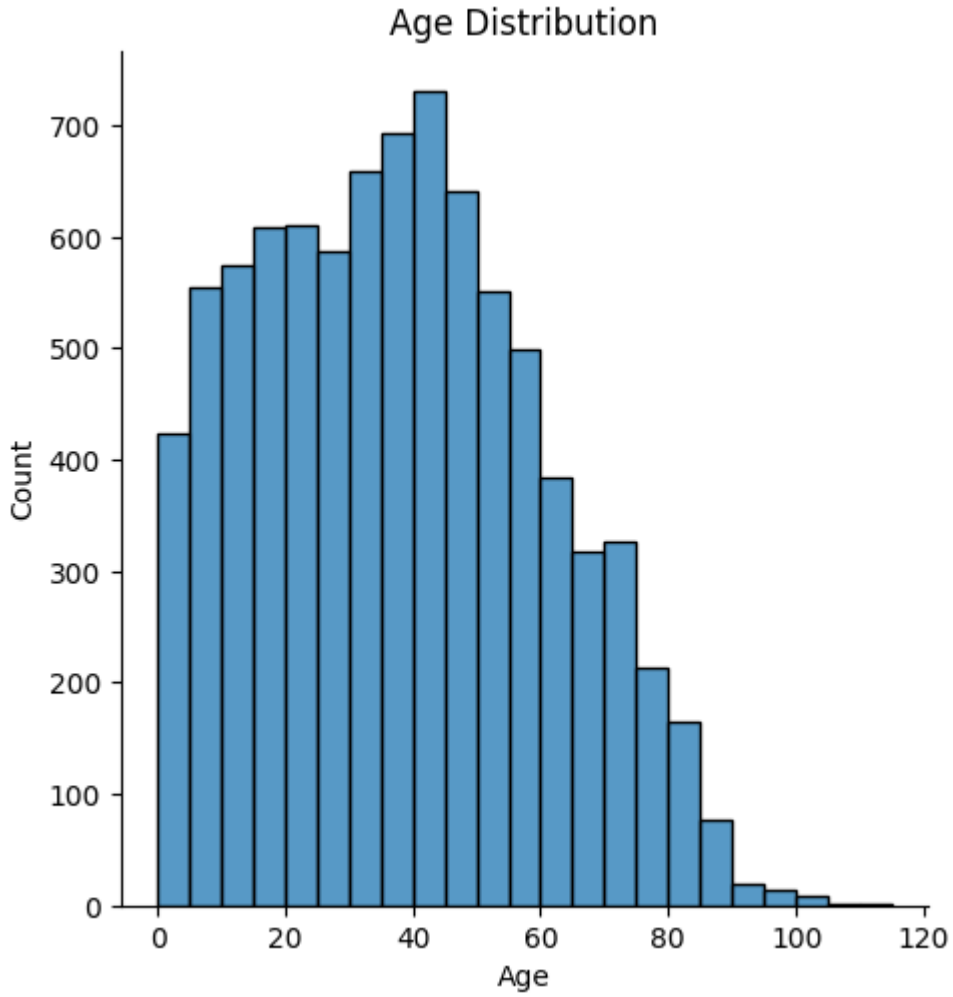

Marital Status by Age
Age
100
80
60
40
20
0
Married
Single
NA
Widowed
Divorced
Marital Status


Unemployment_by_Age
Gender
Female
Male
Count
80
70
60
50
40
30
20
10
0
20
30
40
50
60
70
80
90
Age

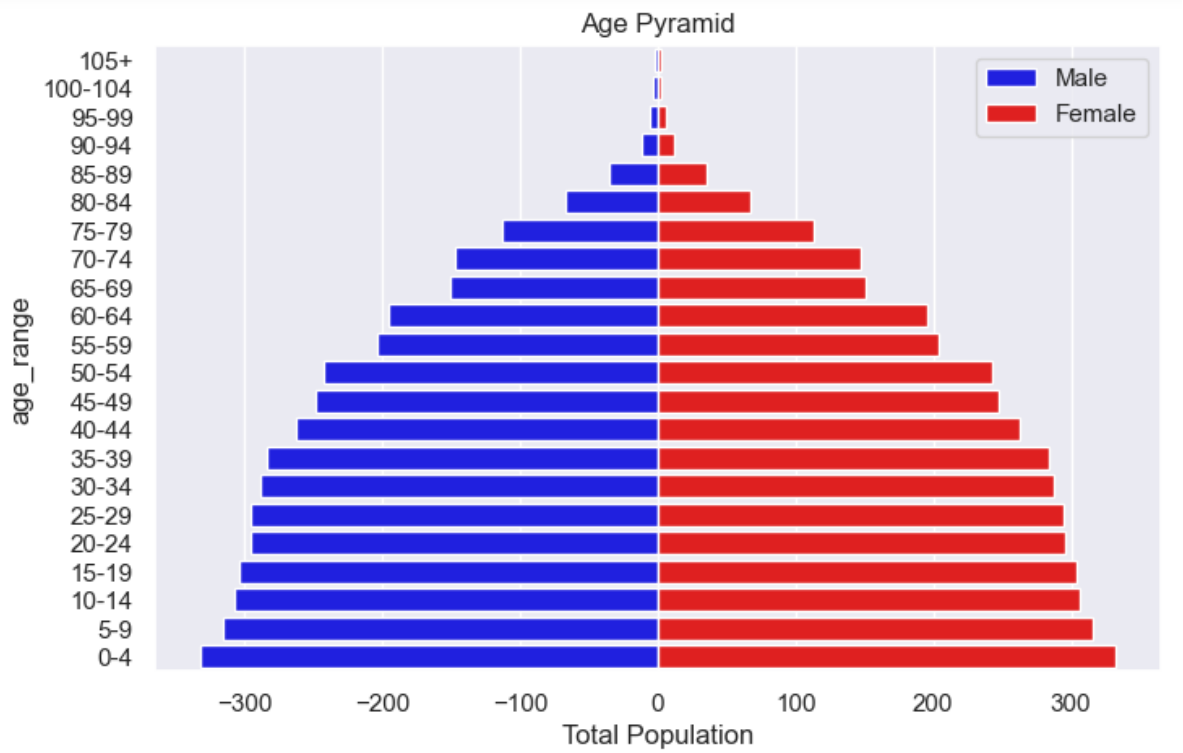
Unemployment was highest among people aged 25 to 50, with women being the most unemployed. Furthermore, those over the age of 65 should not be considered unemployed based on the official retirement age in the UK (GOV.UK); instead, they were classified as 'Retired' for proper analysis.

Age Distribution

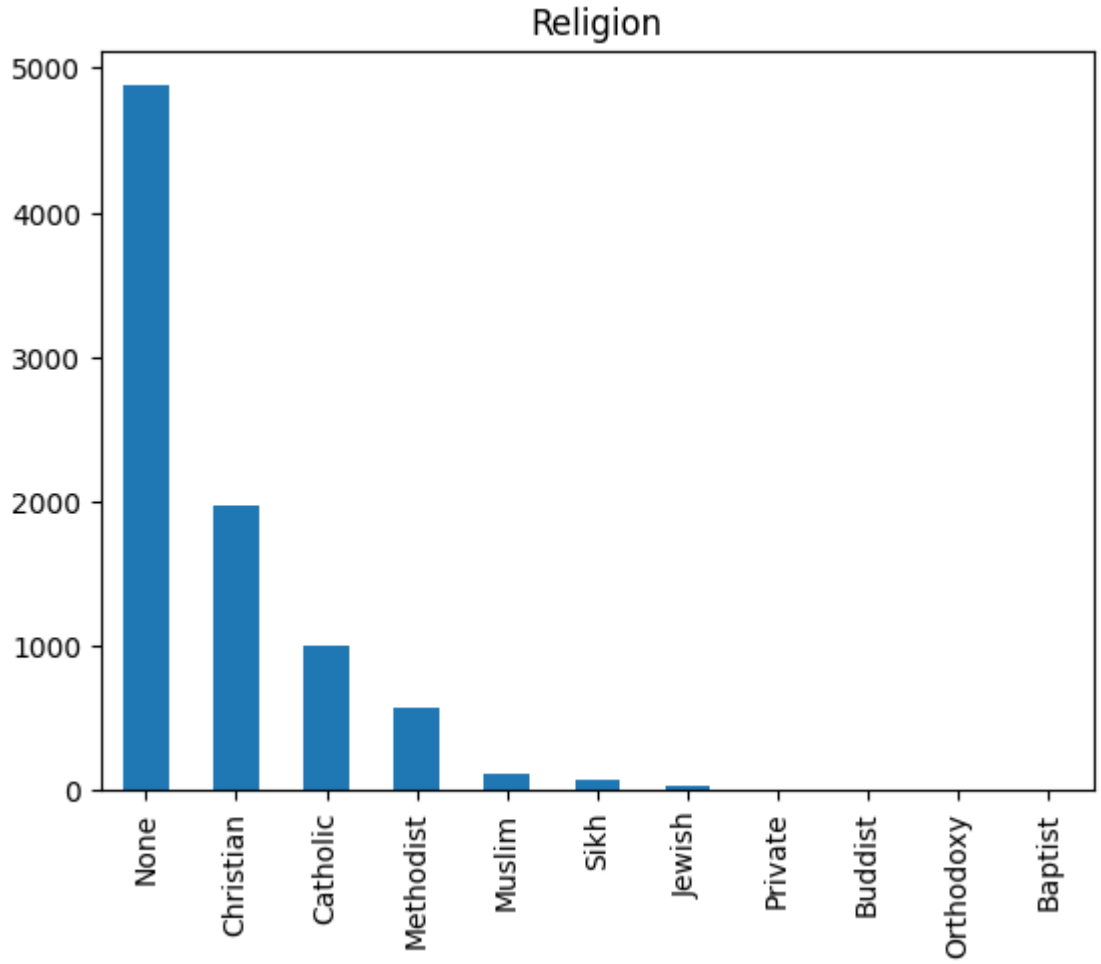
A look at the age distribution reveals a higher proportion of young and middle-aged people. This demonstrates a growing population.

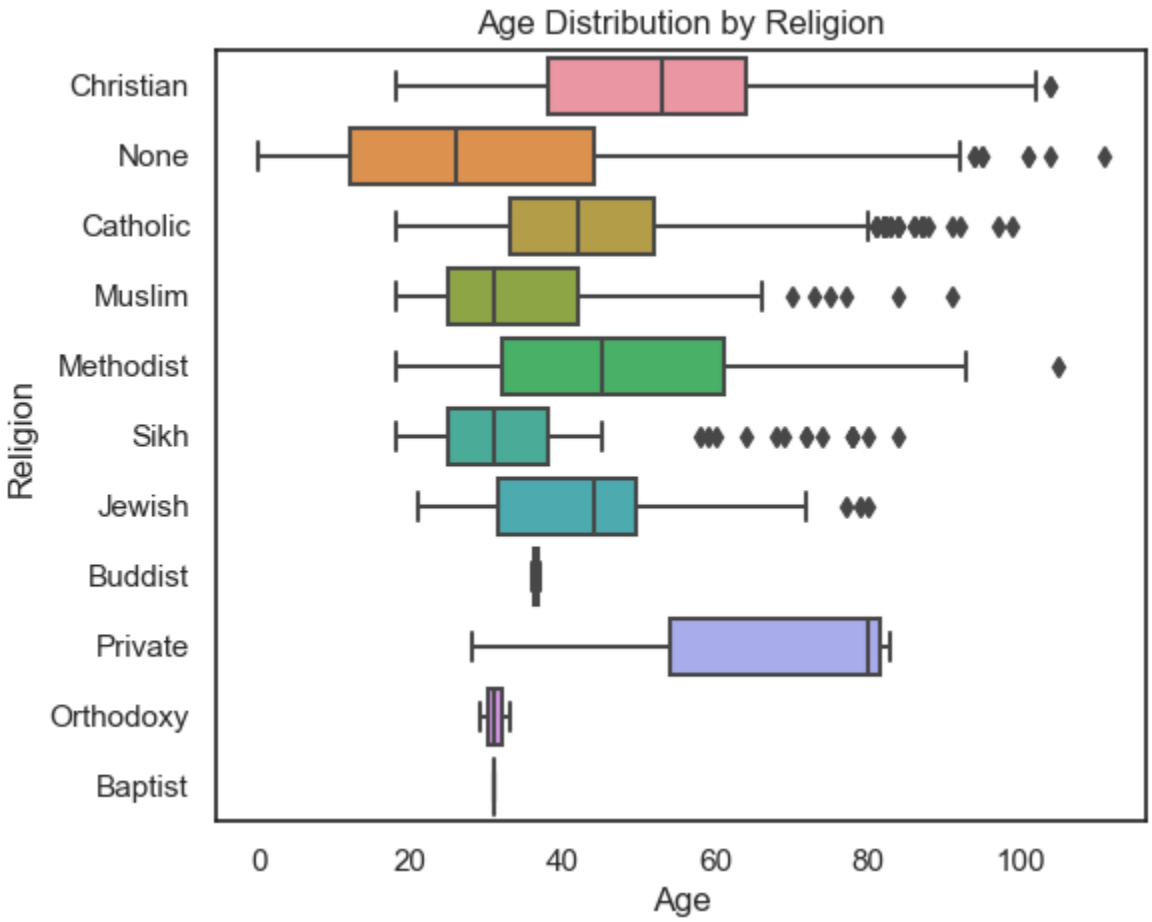


A closer look at the Age Pyramid structure reveals a growing population with a higher proportion of young and middle-aged people. But is a shrinkage at the top, which could indicate a higher death rate.



Religion


Religion
5000
4000
3000
2000
1000
0
None
Christian
Catholic
Methodist
Muslim
Sikh
Jewish
Private
Buddist
Orthodoxy
Baptist



Birth and Death Rate

The number of births and deaths in a population over a given time period is referred to as the birth and death rate. The birth rate is the number of live births per 1,000 people per year in a population, whereas the death rate is the number of deaths per 1,000 people per year in a population. These rates are important indicators of a population's health and stability because they can provide information about the size and growth of a population over time.

From the data, the birth rate is calculated as 8.5 births per thousand, with an evolving birth rate of 229 and 226 per thousand for women of childbearing age between 25 - 29 years and 30 – 34 years, respectively.

	Age	Count
1	105	-2.0
3	102	-3.0
6	98	-1.0
8	96	-4.0
10	94	-1.0
11	93	-3.0
13	91	-1.0
14	90	-5.0
16	88	-10.0
18	86	-5.0
19	85	-9.0
22	82	-13.0
24	80	-9.0
27	77	-4.0
28	76	-6.0
30	74	-14.0
31	73	-14.0
34	70	-7.0
36	68	-25.0

The death rate is calculated as 15.7 deaths per thousand.

It is also worth noting that additional calculations were performed to confirm the population's growth rate, and it was discovered that the population has a growth rate of -0.7163%.

Migration

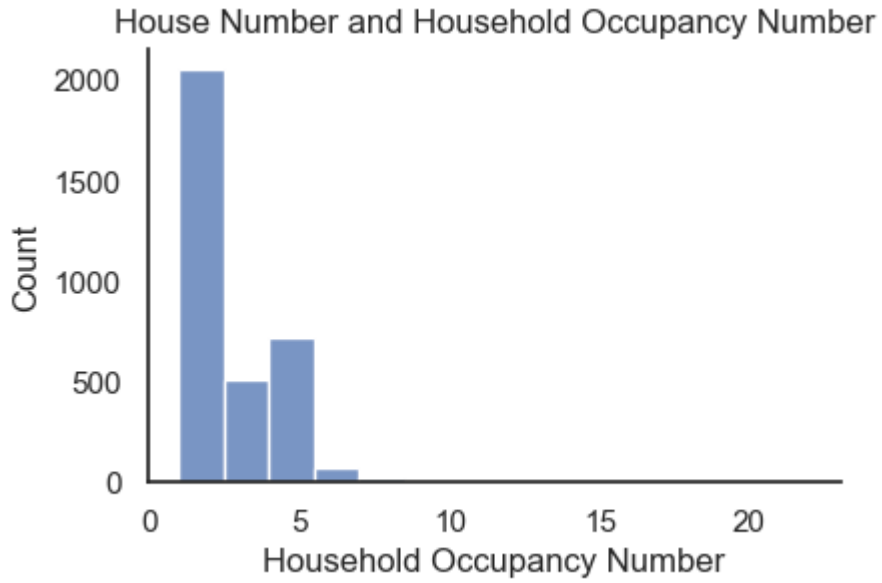
It was discovered that university students account for the majority of the population’s emigration and immigration. In addition, immigration statistics from lodgers and visitors were calculated, and the population has an immigration rate of 22.4 per thousand.

The emigration statistics were calculated using the difference between male and female divorcees, with the population emigrating at a rate of 17.6 per thousand. In general, the population has more immigrants than emigrants.

Household Occupancy Rate

34% of the household is occupied by 1 person, about 26% by 2 people, 15% by 3 people, and 13.7% by 4 people. This shows that some houses are densely populated, and others are not.

	House Number	Household Occupancy Number
mean	42.521077	2.533665
median	25.000000	2.000000
var	2307.198384	3.313068
std	48.033305	1.820184



```
1     34.279859
2     25.790398
3     15.076112
4     13.670960
5      7.435597
6      2.137002
7      0.439110
12     0.292740
10     0.204918
11     0.146370
8      0.117096
9      0.087822
13     0.058548
22     0.058548
20     0.058548
14     0.058548
16     0.029274
17     0.029274
21     0.029274
Name: Household Occupancy Number, dtype: float64
```

There are also 272 lodgers and 21 visitors, totalling a high-density building of 293.

Assumed Commuters and Unemployment

The population has 3606 estimated commuters, accounting for 41.7% of all assumed commuters. Furthermore, the population has 598 unemployed people compared to 4608 employed people. Further analysis reveals that the population has an unemployment rate of 11.5 percent.

The table below breaks down unemployment by age class. It was discovered that the middle-aged class has a high rate of unemployment.

```
40-44    89
35-39    79
45-49    76
30-34    74
25-29    57
50-54    53
55-59    51
60-64    41
65-69    24
20-24    17
70-74    16
75-79     8
85-89     6
80-84     6
15-19     1
Name: age_range, dtype: int64
```

Recommendations

Due to the town's high death rate, there is a need for medical buildings and increased access to healthcare services, including preventive care and chronic illness treatment.

Due to the high number of young and especially middle-aged unemployed in the town, there is also a need to invest in employment and training of people for new skills.

Reference

Humanists UK. (n.d.). Why you shouldn’t write ‘Jedi’ as your religion in the Census. [online] Available at: https://humanists.uk/2021/03/12/why-you-shouldnt-write-jedi-as-your-religion-in-the-census/.

Jedi not a religion, ruling says. (2016). BBC News. [online] 19 Dec. Available at: https://www.bbc.co.uk/news/uk-38368526.

State Pension age timetables. (n.d.). [online] Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/310231/spa-timetable.pdf.

www.simplybusiness.co.uk. (n.d.). What’s the retirement age UK? [online] Available at: https://www.simplybusiness.co.uk/knowledge/articles/2022/02/retirement-age-uk/.

Rune (2022). How to Clean Data using pandas DataFrames. [online] Learn Python with Rune. Available at: https://www.learnpythonwithrune.org/how-to-clean-data-using-pandas-dataframes/.

Sullivan, J. (2018). Data Cleaning with Python and Pandas: Detecting Missing Values. [online] Medium. Available at: https://towardsdatascience.com/data-cleaning-with-python-and-pandas-detecting-missing-values-3e9c6ebcf78b.

matthewbrems (2018). ODSC-missing-data-may-18/Analysis with Missing Data.pdf at master · matthewbrems/ODSC-missing-data-may-18. [online] GitHub. Available at: https://github.com/matthewbrems/ODSC-missing-data-may-18/blob/master/Analysis%20with%20Missing%20Data.pdf.

Legislation.gov.uk. (2020). Marriage Act 1949. [online] Available at: https://www.legislation.gov.uk/ukpga/Geo6/12-13-14/76/contents.