Census Project Report

Introduction

This report is based on the analyses of census data of an imaginary modestly sized town, located in between two larger cities. The purpose of the analyses is to make recommendations for the infrastructural development of an unused plot of land and investment in future services in the town.

Data Cleaning (Raw Data Features)

The census was cleaned in other make it fit for use by correcting the errors and missing values. A comprehensive method (logical, scientific, and statistical) and steps were used to achieve this purpose, and this can be found in the attached Jupyter Notebook. Fig. 1.0 below shows a visual insight into the missing entries of the raw data before cleaning.

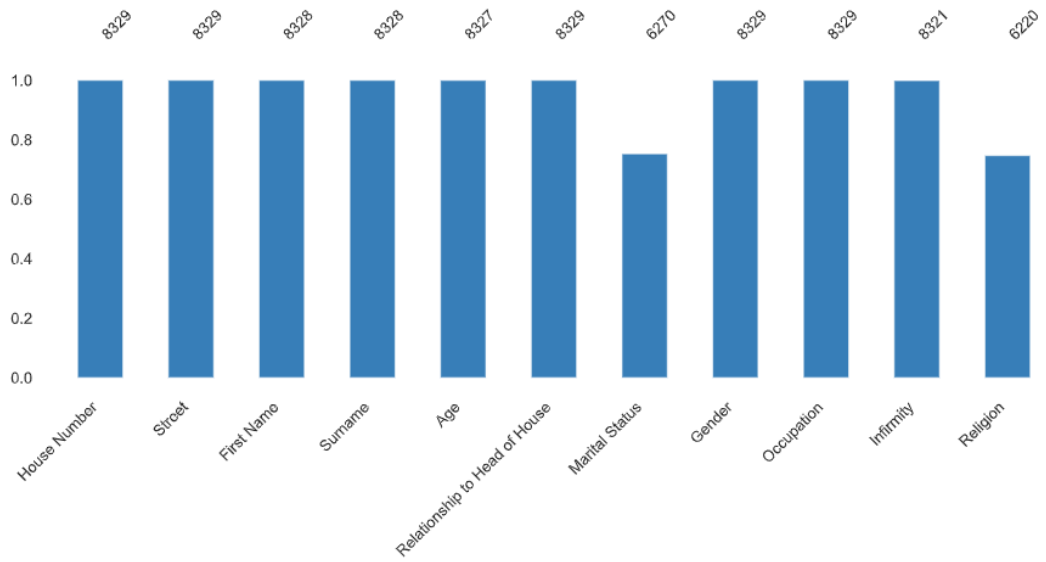


Fig.1.0. Simple visualization of the nullity by column

First, all empty string entries or blank data entries, (“ “), were replaced with NaN so that it can be easily identified in the data set. The census data was cleaned column by column, and entries that could not be imputed with logical information were replaced with “Unknown” like in the case of Infirmity.

Surname and First Name: The two missing values in the first name and Surname were logically imputed by using their house number, street name, and relationship to the head of the house to find their relationship with other members of the household.

Age: Also, the age column which did not have too many missing values, but only two empty string entries, were replaced with the mode age using their relationship to head of house and their occupation. The age entries were converted from object type (strings) to integers, and there were no outliers as all the ages were within the normal life span age range for human beings (Vargas, R.A(2023)). The maximum age was 106, and all the persons with age above 90 years were either retired or unemployed.

Marital Status: Under the Marital Status column, 2059 entries were missing. Most of the NaN values were minors (below 18 years), and according to (GOV.UK), the legal age for marriage is from 18 years and above except for 16 years (on the condition that it is only allowed by the consent of the parents). Hence all individuals with NaN in their marital status who are below 18 years of age were all inputted as "NA" (Not Applicable). The other three NaN values with age above 18 years were inputted using their relationship to other households and the marital status of the other members of the household. The only two outliers that were observed were 16 years and 17 years old females whose husbands were 19 years old, which were given as an exception based on the (UK Marriage Act).

Religion: The religion column had too many NAN values, about 2109 entries, which were mostly individuals below 18 years of age, this was assumed to be the case where the child is allowed to choose their religion as the religion of the parents cannot be legally inferred on the child since by law, since there is freedom of religion or belief (FoRB) (GOV.UK). Even if they were replaced with the parents' religion, it is not certain that the child will retain their parents after 18 years, hence all NaN entries were replaced with Unknown since it can be used to analyse if any religion is growing or not.

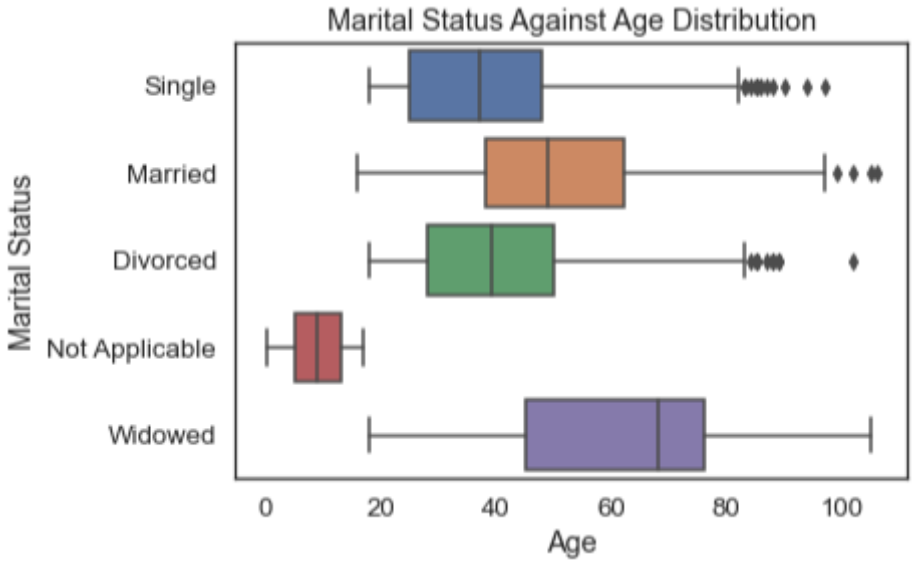
Infirmity: The eight missing entries in the infirmity column were replaced with Unspecified since there are no logical relationships to use in replacing them.

The Occupation column which initially had no issues, was later discovered to have persons who were above 65 years of age, and their occupation was inputted as unemployed. All such cases were replaced with retired, since 65 years is the official pension age according to the United Kingdom law (GOV.UK), and even though is optional to retire at 65 years is likely that individuals who are above 65 years will be unemployable.

Data Insight (Checking for Possible Outliers)

Some plots were made using the data correlations in other to have an idea of the relationships that exist in the census data set.

Fig. 2.0 and Fig. 2.1 below, show the relation between marital status and age. Although there are many outliers as it shows older people are inputted as single, in reality, this may appear to be unlikely true. However, further exploration of marital status likely outliers revealed that most of the people who are 65 years and above, with marital status as single, live alone, and as such is logical to consider them to be single since there are no other household members whose relationship with them proves otherwise. Hence the inputs were not altered. Also, those who are 18 years and are either widowed or divorced were considered to be legally correct or true, since they have reached legal age for marriage. (GOV.UK).



(Fig. 2.0 Age distribution based on marital status.)

	Age							
	count	mean	std	min	25%	50%	75%	max
Marital Status								
Divorced	744.0	40.983871	16.728044	18.0	28.0	39.0	50.0	102.0
Married	2226.0	50.314915	15.961305	16.0	38.0	49.0	62.0	106.0
Not Applicable	2056.0	8.845817	5.053973	0.0	5.0	9.0	13.0	17.0
Single	2943.0	37.691471	14.153906	18.0	25.0	37.0	48.0	97.0
Widowed	360.0	62.061111	20.374943	18.0	45.0	68.0	76.0	105.0

(Fig 2.1. Statistics of age distribution by marital status.)

Approach Used for The Analysis

After cleaning the Census data, Fig. 3.0 below, summarizes the new feature and structure of the data.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8329 entries, 0 to 8328
Data columns (total 11 columns):
 #   Column                        Non-Null Count  Dtype
---  ------                        --------------  -----
 0   House Number                  8329 non-null   int64
 1   Street                        8329 non-null   object
 2   First Name                    8329 non-null   object
 3   Surname                       8329 non-null   object
 4   Age                           8329 non-null   int32
 5   Relationship to Head of House 8329 non-null   object
 6   Marital Status                6270 non-null   object
 7   Gender                        8329 non-null   object
 8   Occupation                    8329 non-null   object
 9   Infirmity                     8321 non-null   object
 10  Religion                      6220 non-null   object
dtypes: int32(1), int64(1), object(9)
memory usage: 683.4+ KB
```

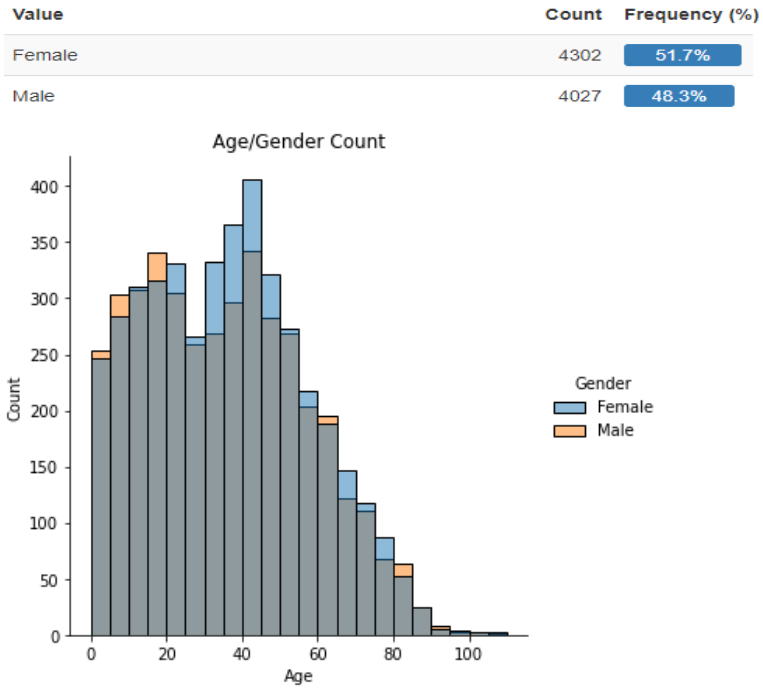
(Table 3.0 feature of the cleaned census data.)

For effective analysis, some new features were added to the data set, this includes:

- **Age Range:** Used for plotting the Age population pyramid with 5 years range (band)
- **Employment Categories (Occupation):** Grouped into six categories: Child, Student, Commuter, Employed, Unemployed, and Retired.

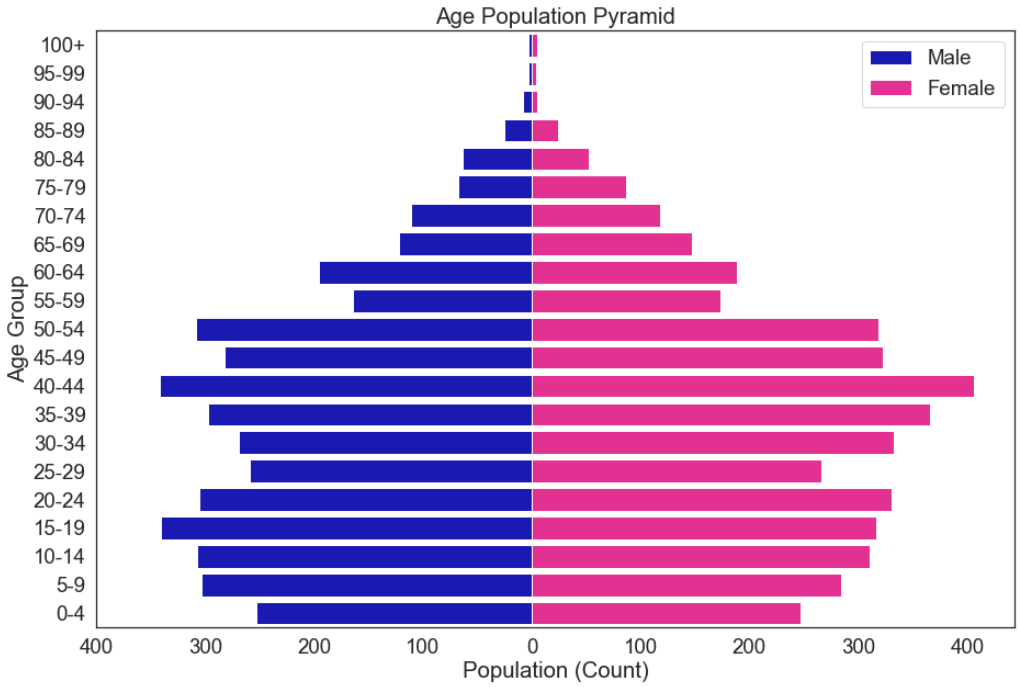
Population Demography

Looking at the age distribution by gender of the data, Fig. 4.0, show that the population of the town consists of a higher number of females within the childbearing age, (20-45), (UK.GOV), and this could mean a high birth rate, but the further detailed analysis will be carried out in other justify this claim by calculating the crude birth rate.



(Fig. 4.0. Population demography based on gender)

Insight from the age distribution pyramid shows that the population demography has a large number of people between the age of 0 – 54 when compared to the older population of 55 – 100+, this interprets a growing population with more young people living in the town. This will further be quantified by calculating the birth rate and death rate. Pre-analysis using different age groups, between 0-4 years indicates sight decline in birth rate when compared to ages of 5-10 years. Then 10-24 years shows a large number of people within the school age (student and university students), while ages between 25-34 years show a decline, this suggests that most of the people between the school-age leave the town after graduation from the university (especially males). From 35-54, years also suggest a larger population of the work force but can only be justified by the actual number of employed individuals within this age group.



(Fig. 4.0. Age distribution pyramid)

Data Analysis

Infirmity

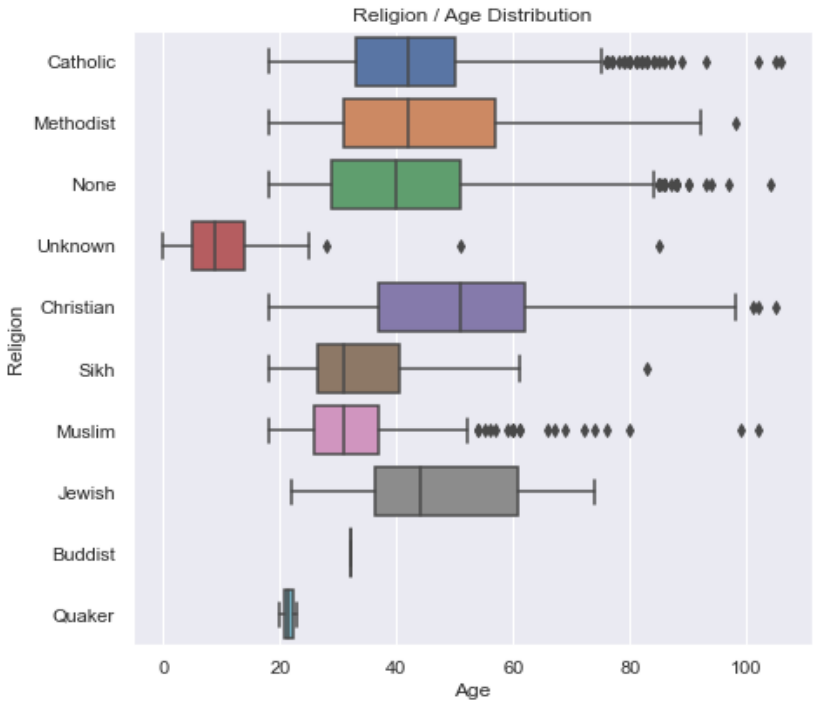
The percentage of people with infirmity in the town is very small (less than 1%) and this can be considered insignificant when compared to the entire population of the town as can be seen in Fig 5.0 below. Based on this observation building a hospital or clinic will not have commiserate social-economic value to the town as to other infrastructural needs of the town.

Value	Count	Frequency (%)
None	8261	99.2%
Physical Disability	17	0.2%
Mental Disability	13	0.2%
Blind	10	0.1%
Unknown Infection	8	0.1%
Unspicified	8	0.1%
Deaf	8	0.1%
Disabled	4	< 0.1%

(Fig 5.0. Statistics profile of Infirmity)

Religious Affiliation

In Fig 6.0, Some of the religious groups with very smaller members, (Sikh, Muslim, Buddist) appear to be growing as the median age of members is low but the number is so small to be considered for analysis and decision-making.




```
Religion : Median
Christian: 51.0
Catholic : 42.0
Methodist: 42.0
Unknown:   9.0
None:      40.0
Sikh:      31.0
Mulsim:    31.0
Jewish:    44.0
Buddist:   32.0
Quaker:    21.5
```

(Fig. 6.0 Age distribution by religious affiliation and Religion median by age)

Even though there are more Christian members in the town (21.4%), their median age seems to consist more of older people this means the religion is not growing rapidly and will not have larger members in the future.

In general, considering the religious diversity (4 major religious groups) of the town and a large number of people who have no religion, (33.6%) or whose religion is unknown, (25.3%), building a religious building, (church or mosque) will not be of benefit to the majority of people living in the town.

Common Values

Value	Count	Frequency (%)
None	2800	33.6%
Unknown	2109	25.3%
Christian	1781	21.4%
Catholic	881	10.6%
Methodist	547	6.6%
Muslim	115	1.4%
Sikh	55	0.7%
Jewish	38	0.5%
Quaker	2	< 0.1%

(Fig 6.1. Statistics profile of Religion)

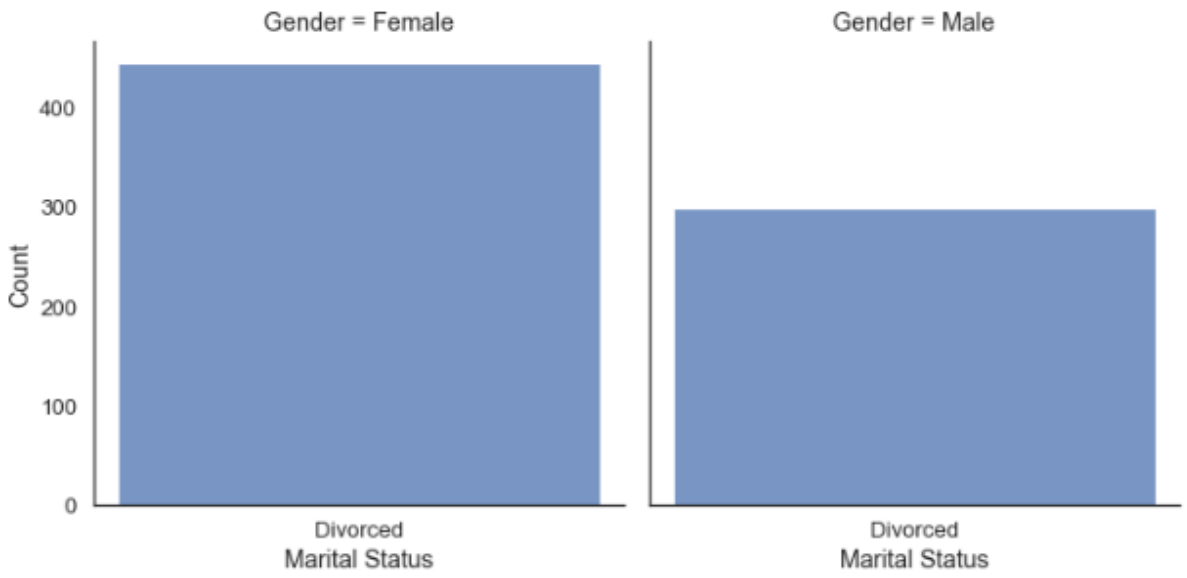
Marital Status: (Marriage and Divorce)

There are 744 divorcees living in the town which is about 8.9% of the town's population.

Value	Count	Frequency (%)
Single	2943	35.3%
Married	2226	26.7%
Not Applicable	2056	24.7%
Divorced	744	8.9%
Widowed	360	4.3%

(Fig.7.0. statistics profile of Marital status)

There are more divorced females than males in the town with a total of divorced females 445 and 299 divorced males, as seen in Fig. 7.1. This suggests that most of the males leave the town after they are divorced. Concerning (Fig 2.0), most people marry and divorce at a young age, as observed by their mean age this forms part of the reason why the population tends to shrink between males of 25-39 years age group as seen in the age distribution pyramid as most male divorcees tend to leave the town after getting a divorce (fig 4.0). This will be considered for emigrants.



(Fig. 7.1. Divorced relationship with gender)

Birth and Death Rate

To calculate the crude birth and death rate some assumptions were made. (This includes:

- ❖ Considering the number of children per year (using 5 years interval)
- ❖ For death rate age between 60 - 100+ years were considered.

rude birth Rate $= (\frac{\text{number of children within a year}}{\text{Total popultion}})$ *1000

Crude Death Rate $= ((\frac{Death\ within\ 5\ years}{\text{Total population}}) * 1000)/5$

Years	Annual Crude Birth Rate per 1000 population (base on age 0 to 9)	Years Range	Annual Death Rate per 1000 population (based on ages 60-100+
Current years	11	(60 – 69)	3
5 years ago	12	(70 – 79)	5
10 years ago	14	(80+)	5

(Table.1 Evolving crude birth rate.)

Table.1.0 above shows the calculated crude annual birth and death rate for different years. Even though the birth rate appears to have slightly decreased by 1 child per 1000 of the population, within five years and 3 Children per 1000 ten years ago, the total population is still considered to be growing since the annual crude birth is still higher than the annual death rate at different years.

Also, the constant value of death rate between 70 – 79 years and 80+ suggests that the town has high longevity as they grow older. This could mean there is a need for old age care homes.

Migration (Immigrants and Emigrants)

University and Ph.D. students make up the majority of immigrants and emigrants in the town because they leave after completing their studies. This may help to explain why the population between the ages of 15 - 19 has increased significantly while that of ages between 25 - 29 has slightly decreased as observed from the age distribution pyramid above.

True Immigrants are considered to be those who have moved into the town, these are visitors and lodgers who are not divorced which is a total of 573 individuals.

Occupation and Commuters

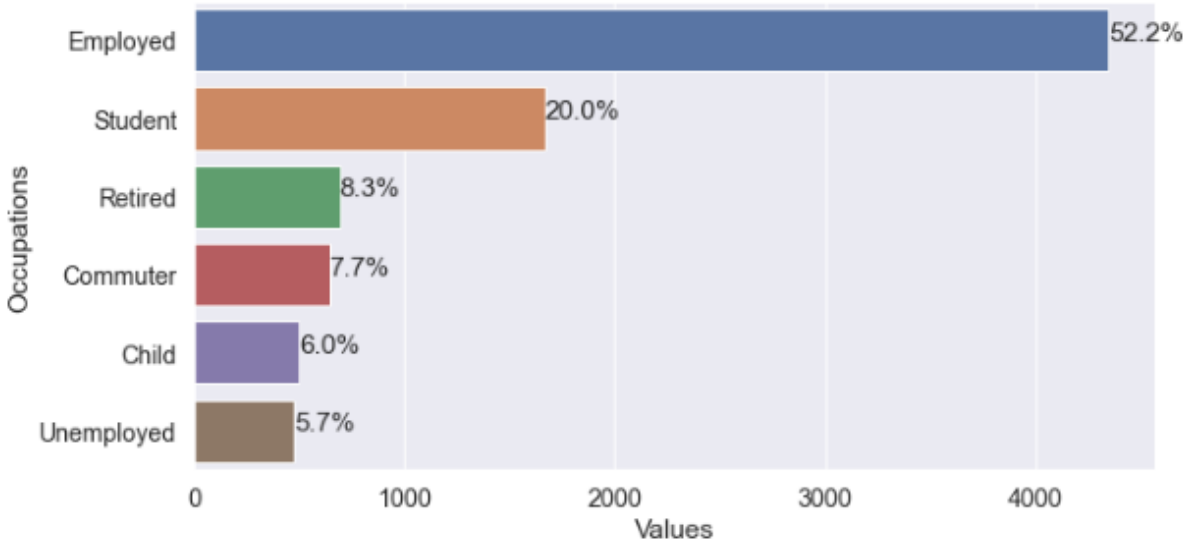
There is a significantly large percentage, (51%) of the population who are employed as compared to the percentage of people who are either unemployed (5.7%) or retired (8.3%), as shown in Fig 8.1 below. However, the general observation of the age distribution of the unemployed individuals is 41.82 years, which is relatively close to the mean age of the employed individuals at 42.86 years. This suggests that there is no significant difference in the age distribution between employed and unemployed individuals.

Furthermore, younger age groups below 41 years can be said to be likely to be unemployed individuals since that is the median age of unemployed individuals and is slightly lower than that of the employed individuals at 42 years (Fig. 8.0), so there might be need to invest in training younger people for new skills, but this will be based on other pressing priorities of the town.

	Age							
	count	mean	std	min	25%	50%	75%	max
Occupation								
Child	500.0	2.098000	1.377022	0.0	1.00	2.0	3.0	4.0
Commuter	672.0	23.755952	9.114275	18.0	19.00	21.0	22.0	65.0
Employed	4319.0	42.864089	12.025614	19.0	33.00	42.0	52.0	71.0
Retired	693.0	76.089466	7.034849	66.0	70.00	74.0	80.0	106.0
Student	1669.0	11.485321	3.991064	5.0	8.00	11.0	15.0	18.0
Unemployed	476.0	41.819328	10.803748	22.0	33.75	41.0	50.0	65.0

(Fig. 8.0. Statistics of Occupation distribute by age.)

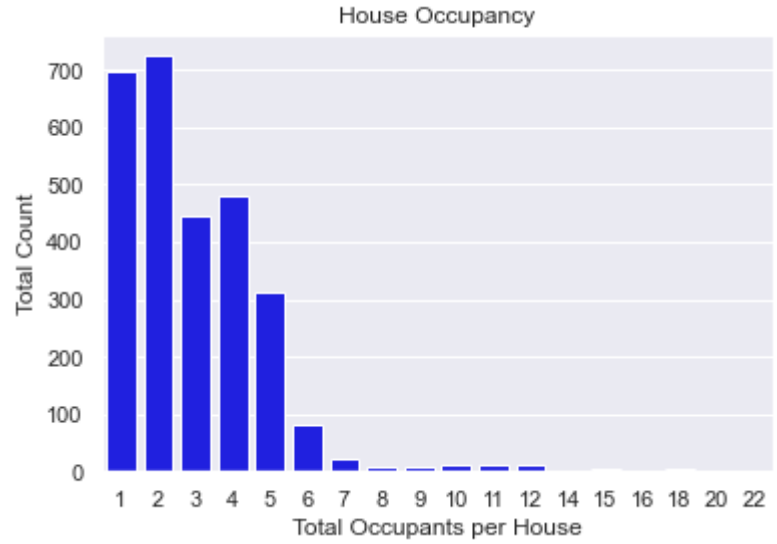
Since the town has no university, all the people living in the town whose occupation descriptions include, university student, Ph.D. student, lecturer, Academic, professor, also all other occupation descriptions relating to train and airline workers, since the town has no train station are all considered to be commuters. This group were classified as (Commuter), with total of 7.7% of the entire population of the town as shown in fig 8.1 below,



(Fig. 8.1. Statistics Profile of Occupation)

Occupancy Rate

There are more houses with only 1 and 2 occupants, which is a total of 697 and 725 houses respectively as seen in Fig. 9.0. With reference from the 1930s until the 1980s, the number of bedrooms in the United Kingdom, fluctuated between 3.16 and 3.63 rooms per house, however declining family sizes caused this number to gradually decrease from the 1980s onwards, and in the 2010s the average number of bedrooms has fallen to 2.95 rooms per house(Statistica,2022). The general United Kingdom house occupancy standard is, no of persons per room is; 1 room per 2 persons, 2 rooms per 3 persons, 3 rooms per 5 persons, and 4 rooms per 7 persons, as such this means that the houses are not overused.



	Household Occupancy	Counts
1	1	697
0	2	725
3	3	446
2	4	479
4	5	313
5	6	80
6	7	21
10	8	9
11	9	9
8	10	10
9	11	10
7	12	12
17	14	1
12	15	3
15	16	2
13	18	3
14	20	2
16	22	2

(Fig. 9.0. Total number of occupants per household)

Summary

Parameters	Total Number per1000 Population	Percentage per Population (%)
Crude birth rate	11	**1.1**
Crude death rate	3	**0.3**
Immigrants	41	**4**
Emigrants	18	**1.8**
Commuters	77	**7.7**
Employable population (16 - 65yrs)	687	**68.8**
Unemployed population	57	**5.7**
Student	200	**20**

(Table 2.0. Summary of Results)

Recommendations

Considering, the large population of school-age students, who will in the nearest future be potential commuters (University Students), and the total percentage of commuters at 8.1% of the population, so there is a need to build a train station. This might be capital intensive but, in the near future, will also increase revenue generation and provide more employment in the town.

Also, there need to invest in vocational skill acquisition, considering that 5.7% of the population is unemployed. Since not all school-age students will attain a university education a proportion of them (4.3%) of them will require employability skills. This is chosen over investing in old age care facilities (75 years and above) since the of old people that may require care is 4.15 of the population and there are already care homes in the town.

While there is a modest influx of immigrants (4%) into the town, constructing low-density housing may not be advantageous for individuals who rent out rooms to lodgers, and if the houses are not fully utilized, it may not result in economic benefits for the town.

References

Bradford, S. (2023) *User guide to birth statistics 19 January 2023*. Available online: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/livebirths/methodologies/userguidetobirthstatistics/pdf [Accessed 27/April/2023]

Freedom of religion or belief: understanding this human right (2022) GOV.UK. Available online: https://www.gov.uk/guidance/freedom-of-religion-or-belief-understanding-this-human-right. [Access 24/04/2023]

Getting married (no date) Citizens Advice.
Available online: https://www.citizensadvice.org.uk/family/living-together-marriage-and-civil-partnership/getting-married/#:~:text=All%20couples%20can%20marry%20if. [Accessed 24/April/2023].

Gov.uk (2011) *Marriages and civil partnerships in the UK GOV.UK*. Available online: https://www.gov.uk/marriages-civil-partnerships. [Accessed 01/04/2023]

Gov.uk (No Date) Child Employment Age.
Available Online: https://www.gov.uk/child-employment [Accessed 10/04/2023].

Statistica.com (2022) Average number of bedrooms in new British houses 1930-2020 Statista. Aaron O'Neill. Available online: https://www.statista.com/statistics/1056041/average-number-bedrooms-new-british-houses-1930-2020/#:~:text=From%20the%201930s%20until%20the [Accessed 27/04/2023]

Vargas, R.A. (2023) World's oldest person on the keys to longevity: 'stay away from toxic people'. *The Guardian*, 26 January.
Available online: https://www.theguardian.com/us-news/2023/jan/26/worlds-oldest-person-115-maria-branyas-morera-california#:~:text=Mar%C3%ADa%20Branyas%20Morera%20has%20lived
[Accessed 24/April/2023].