

Census Project Report

This report examines the census data of a town to gain insight into key demographic factors. The insight gained is then used to inform the decision-making process on what to build on an unoccupied expanse of land and where to invest Government funds.

The first section of this report recounts the cleaning of the data. Subsequent sections detail the analysis done on the clean data, looking closely at factors such as birth rate, death rate, marriage, divorce, employment and housing.

Data Cleaning

The census data was cleaned to address discrepancies and missing values. Step by step details of the cleaning done can be found in the python notebook file attached to this report. A summary of the cleaning done on the data set is as follows:

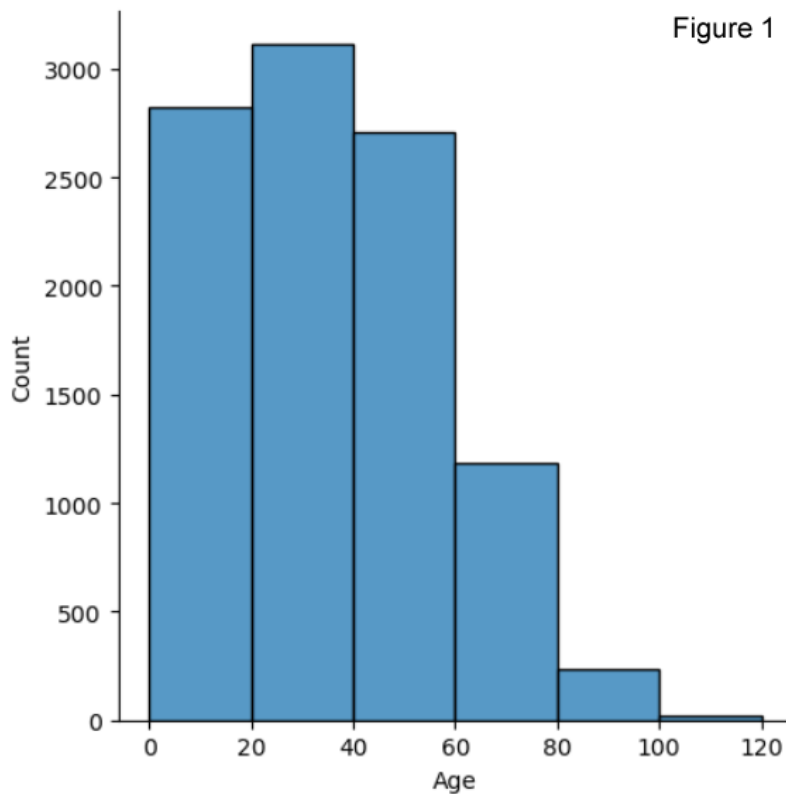
- One duplicate row was located and dropped from the entire dataframe.
- The values in the Age column were rounded down and cast from floating point numbers to integers.
- A blank entry in the Marital Status column was investigated by looking at all individuals listed at the same physical address as the individual with the blank space entry. Doing so provided a clear indication of the actual marital status of this individual. The status was then changed accordingly.
- An individual aged 17 was listed as Head of household. Investigations revealed that this individual, a female was married to a Male aged 20 and living in the same household. A newborn child was also observed in this household. The female was below the legal marriage age of 18 but above 16 which is the legal age for marriage with parental consent therefore no changes were made to her marital status. However, the relationship to head of house for the female was changed from 'Head' to 'Wife', considering that all heads of household should be 18 or older. The relationship to head of house for the spouse was also changed from 'Husband' to 'Head'.
- Two blank entries in the Relationship to head of house column were investigated individually and changed to the correct values of 'Daughter' and 'Husband' respectively.
- 34 rows with 'Visitor' as entries in the Relationship to head of house column were dropped from the entire dataframe, reasoning that these individuals are likely to be from out of town and therefore do not actually contribute to the demographic makeup of the town. Failure to exclude these individuals could adulterate any analysis done.
- A row with a blank entry in Gender column was investigated. Looking at the first name of the individual it was clear that the gender should be indicated as 'Male' and this change was implemented.

- Three rows with blank entries in the surname column were located. These rows were further investigated by looking at the other individuals in the census data listed at the same physical address. This gave a clear indication of what the surnames should be in each case and the changes were implemented accordingly.
- A blank entry in the infirmity column was changed to indicate 'None' instead.
- In the Religion column, 'Methodist' and 'Catholic' entries were changed to 'Christian'. 'None', 'Nope' and 'Agnostic' entries were changed to 'Not Religious'. Blank, misleading and nan entries were first changed to 'Unknown'. I observed that people living at the same physical address mostly practiced the same religion therefore these 'Unknown' values in the religion column were imputed with the highest occurring religion of the Heads of household listed at the same house number and street as the individuals with the unknown status. After this operation, only 36 individuals were left with an unknown religious status.
- Four entries aged 18 and Three entries aged 19 with 'Widowed' in the marital status were observed. Since death can occur at any age, these statuses were left as 'Widowed'.
- 'Divorced' entries in the marital status column for individuals Aged 18 – 20 were changed to 'Single' since the legal marriage age is 18 and by law, divorce applications can only be filed after one year of marriage, with the settlement process itself taking at least another one year after filing (Wells Burcombe Solicitors, 2020 and Wooley & Co Solicitors, 2022).

The overview of the clean data is shown below.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10083 entries, 0 to 10082
Data columns (total 11 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   House Number                          10083 non-null  int32
1   Street                                10083 non-null  object
2   First Name                            10083 non-null  object
3   Surname                               10083 non-null  object
4   Age                                   10083 non-null  int32
5   Relationship to Head of House         10083 non-null  object
6   Marital Status                        7600 non-null   object
7   Gender                                10083 non-null  object
8   Occupation                            10083 non-null  object
9   Infirmity                             10083 non-null  object
10  Religion                              10083 non-null  object
dtypes: int32(2), object(9)
memory usage: 787.9+ KB
```

Birth Rate Analysis



Observing the displot done on the ages, the age band 0-20 was seen to be the only band with a peak lower than the adjoining age band (20-40) suggesting a birth rate reduction over the last 20 years. Some further statistical computation was done to confirm this.

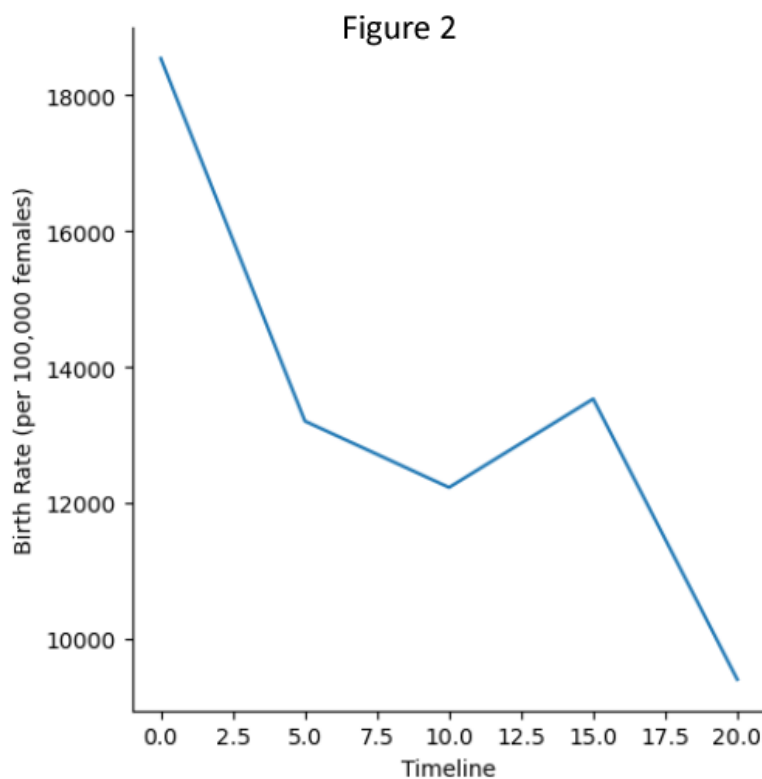
The current birth rate was estimated by computing all recent births which are the individuals between the ages of 0 and 1. Females in the age range for childbearing (18 – 45) were also computed (ONS, 2022). The birth rate was calculated as follows:

Birth rate (per 100,000 females) = Recent births * 100,000 / Total females in childbearing age band

To estimate previous birth rates, this same computation was performed, graduating the age bands considered for recent births and childbearing in steps of 5 years. Care was taken to stop this analysis at the age of 65 for mothers as this age is very close to the average life expectancy and many mothers above that age bracket could already be deceased and absent from the census data (ONS, 2021).

The outcome of the Birth Rate analysis is shown below:

	Birth Rate (per 100,000 females)	Years Ago
0	9397.387273	0
1	13528.896673	5
2	12225.153914	10
3	13202.247191	15
4	18542.336549	20



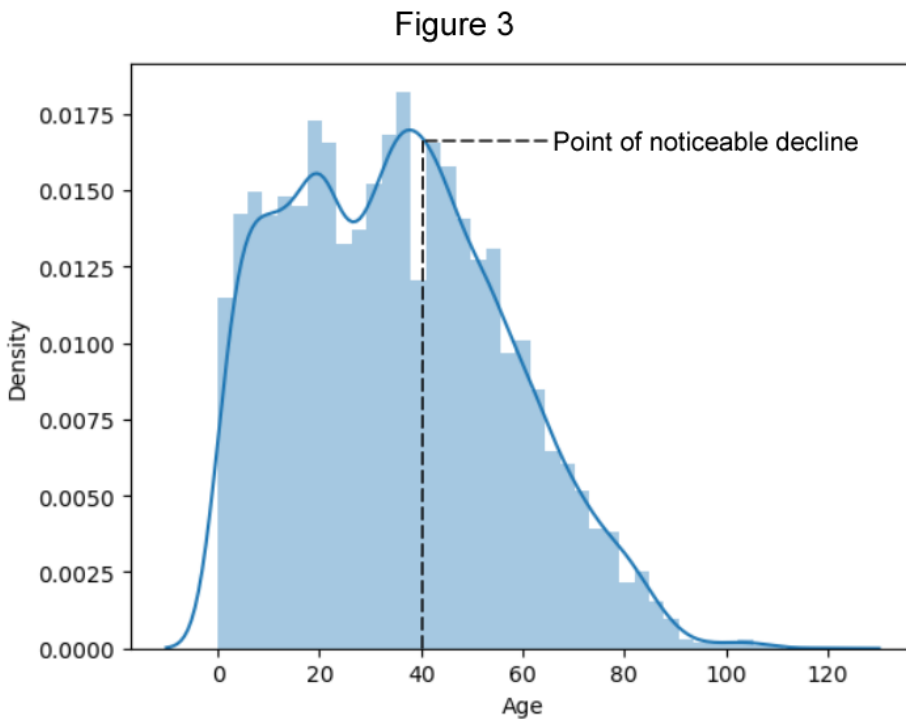
Birth rates have dropped along every 5-year interval in the last 20 years, except for a minor increase in birth rates from 10 years ago to 5 years ago (as shown in Figure 2). There has been an overall decrease in birth rates over the last 20 years from 18,542 to 9,397 per 100,000 females of childbearing age.

For the sake of further analysis, the crude birth rate (WHO, 2023) was also computed as:

Crude Birth rate (per 100 thousand of total population) = Recent births * 100,000/ Total Population

This gave a current crude birth rate of 2,211 per 100,000 of the total population annually.

Death Rate Analysis



The crude death rate was calculated as the number of deaths in a given period divided by the population exposed to risk of death in that period (WHO, 2023). With no death records, this analysis was done by comparing specific age brackets to their adjoining age brackets, computing the difference in the counts of individuals in both age brackets. This gives an idea of the number of deaths as individuals graduate from one age band to the next. This analysis assumes that the number of individuals in each of these age bands should be the same as when they were in a reference age band (the lower age band). This analysis also assumes that the total population x years ago is roughly the same as now, where x is the number of years being used as the gap. Clearly this analysis is flawed but the inaccuracy can be minimized by using smaller gaps between the age brackets. I used a 5-year gap and started from the age of 40 which is when there is a noticeable decline in the population count as shown in the age distplot (Figure 3). This assumes that all deaths which occur below the age of 40 are negligible when put in juxtaposition with deaths that occur after 40.

Using this system, the crude death rate was calculated to be 1,620 deaths per 100,000 of the total population annually.

Population Growth

Over the years there has been a steady decline in the birth rate. However, this decline does not suggest an outright population decrease. The crude birth rate is higher than the crude death rate and the overall effect of this is a gradual increase in population. If the birth rate continues to drop and the death rate (which is less likely to change) remains within a close range of the current death rate, the population growth margin would become smaller every year until eventually the crude death rate would supersede the crude birth rate and then the population would begin to decrease in size.

It can be inferred that the population of the town is expanding currently, however, considering that the crude birth rates and death rates were evaluated per 100,000 and that the total population of the town is just 10,083, this population growth is not significant enough to inform any key decision-making process. To put this into context, the difference between the current birth rate and death rate translates to the town's population increasing by a value of just 59 in the last one year.

Employment and Education

Figure 4



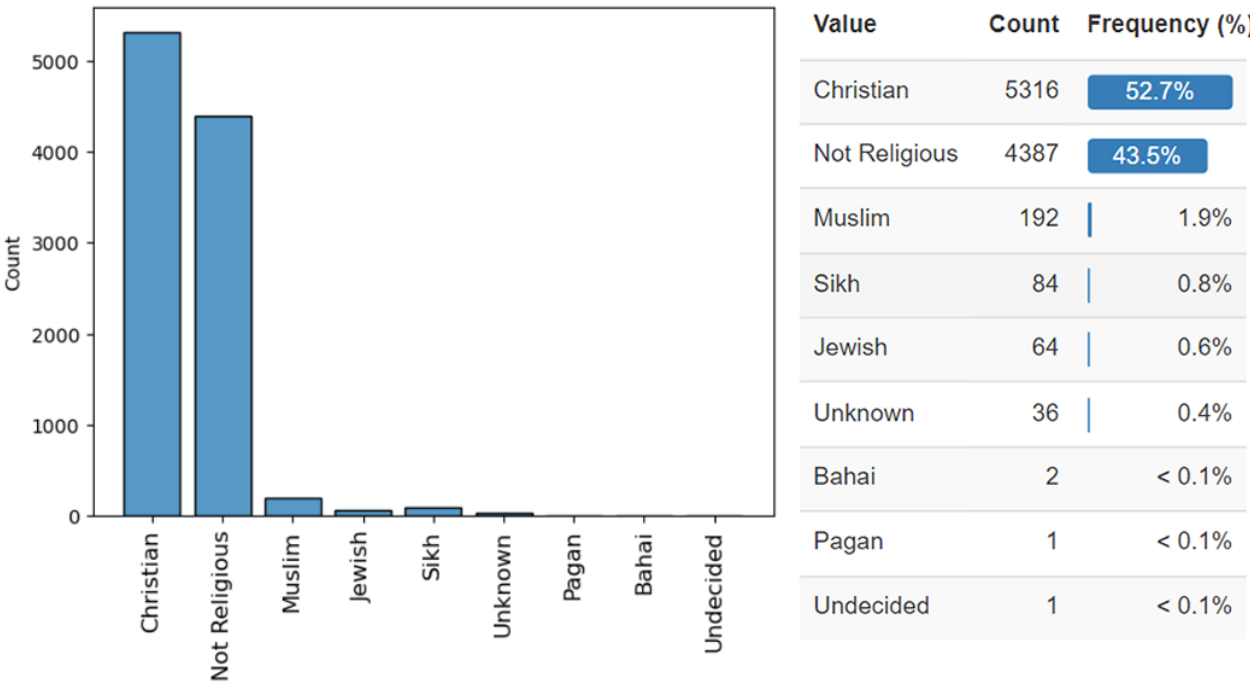
The employment rates across the various age bands and the number of individuals in various categories of employment were evaluated. The number of university students was also computed to gain clearer insight into the commuting requirements of the town.

The retirement age was observed from the dataset to be 67 years. The employment rate analysis was done by considering three age groups (OECD, 2023):

- Ages 16-24 (New to labor market): The employment rate was calculated to be 21.25%. However, 77.26% of the individuals in this age group are students. Considering only the non-students in this age group, the employment rate was calculated to be 93.44%.
- Ages 25-54 (Prime of working): The employment rate was calculated to be 89.69%.
- Ages 55-67 (Approaching Retirement): The employment rate was calculated to be 89.95%.

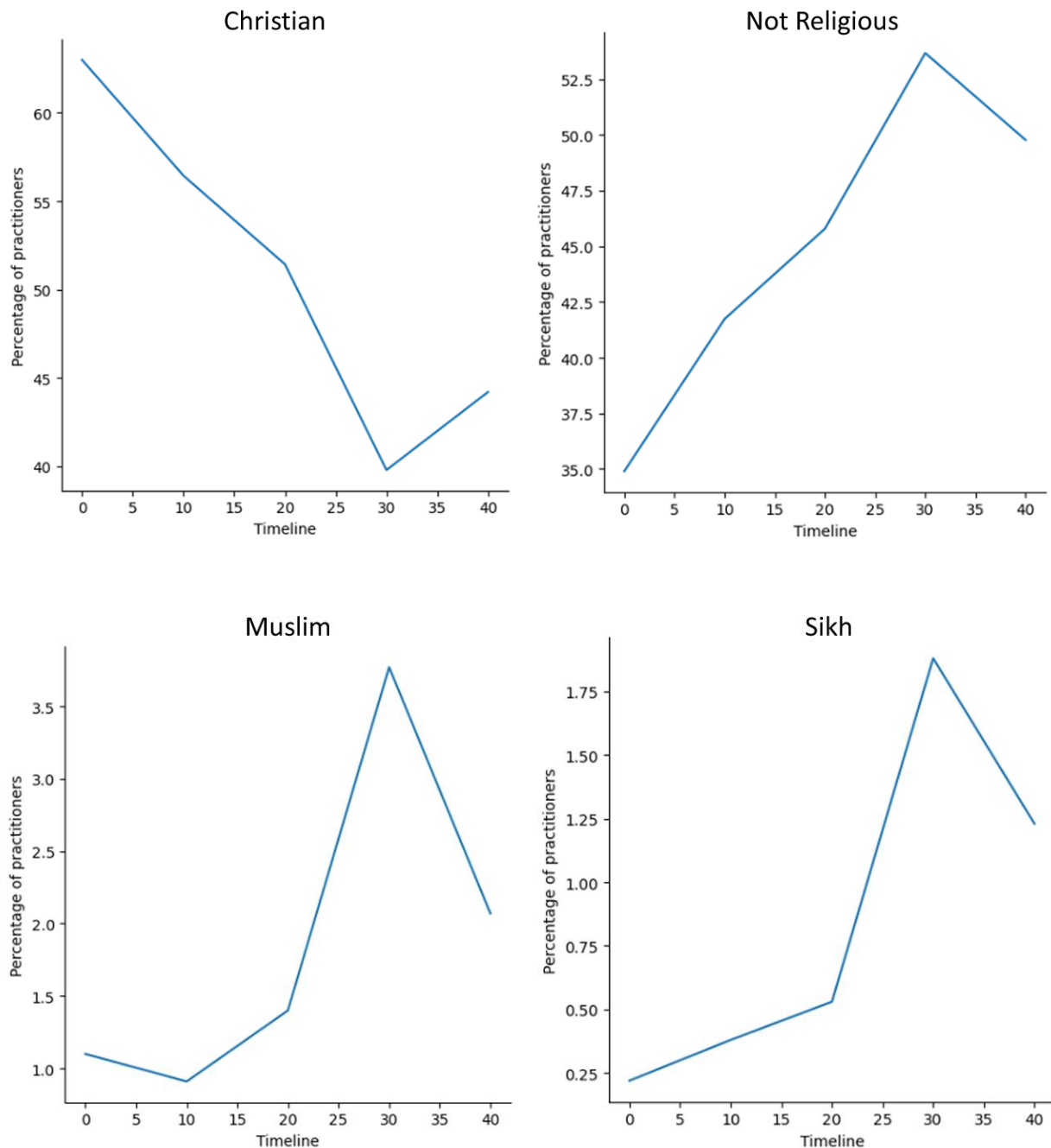
The displot in Figure 1 shows that the age bands 20-40 and 40-60 make up a large portion of the total population. The population pyramid (Figure 4) also highlights this distribution. The high employment rates across these age bands indicate a large number of employed individuals. This number of employed individuals was calculated to be 5,390. The number of university students was also calculated to be 1,111. Considering that the university students need to commute to universities in neighboring towns and that the employed individuals also require to commute to their places of employment, it is apparent that a means to safely transport these individuals (which make up more than half of the total population) should be a top priority.

Religion



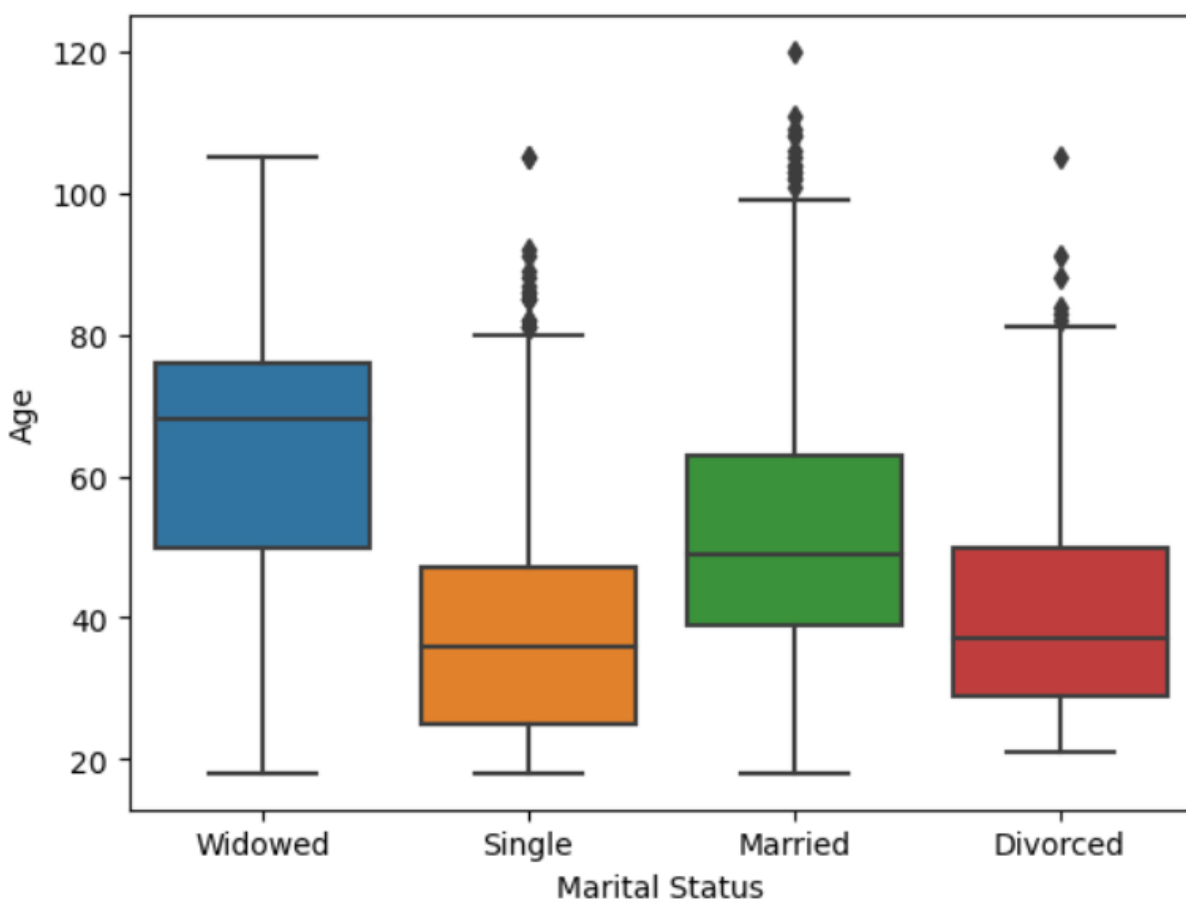
Christians make up 52.7% of the total population while unreligious people constitute another 43.5%. The remaining 3.8% consists of other religions, undecided folks and those with an unknown religious status.

The recent trends in the statistically significant religions were studied by comparing the percentage of practitioners for each religion across several age bands above the age of 16. The individuals aged below 16 were not considered in this analysis as these individuals are still in the spiritual and mental formative stage and most likely have not developed a strong opinion on religion yet.



Christianity was observed to be getting less popular, dropping from 62.99% to 44.21% between the age bands 56-65 and 16-25. The Jewish Religion was also seen to be on the decline. Having no religion increased in prevalence from 34.91% to 49.77%. Muslim and Sikh religions showed rapid increase in popularity with Muslim religion increasing from 1.10% to 2.07% (an 88% increase) and Sikh Religion increasing from 0.22% to 1.23% (a 459% increase). Despite these growths, the total number of practitioners for both religions remain very low relative to the general population therefore these religions would not inform any key decision-making exercise.

Marriage and Divorce



With 18 years being the legal marriage age, only the individuals aged 18 and above were considered for the analysis on marriage. A new dataframe containing only individuals in this age bracket was created for this purpose.

The probability of marriage was computed by summing up the number of married, divorced and widowed individuals and then dividing this sum by the number of individuals with a known

marital status. Divorced and Widowed individuals were included in this calculation because they too were once married.

The probability of marriage was found to be 50.25% indicating that 50.25% of the adults in the town get married at least once in their lifetime.

The probability of divorce was considered next. The 25th percentile age of divorced individuals was computed to be 29 years (as shown in the boxplot) which means most of the divorces were finalized at the age of 29 or after. Only individuals aged 29 and above were therefore considered for the analysis on divorce. Considering a wider age bracket which includes younger folks would have produced less accurate results as many of the younger individuals listed as married could be divorced in future years. My analysis factored this in by allowing some time for those divorces to occur. This logic was further strengthened by divorce statistics which show that most divorces are finalized after the first 10 years of marriage (Harbour Family Law, 2022).

The probability of divorce was then calculated by dividing the number of individuals with a divorced status in the age band being considered by the sum of this same number of individuals and the number of individuals still listed as married.

The probability of divorce was found to be 19.40%.

Housing Analysis

A new data frame was created with an 'Address' column which combines the values in the 'House Number' and 'Street' Columns of the original dataframe. The first 10 rows of this new dataframe are shown below:

	House Number	Street	First Name	Surname	Age	Relationship to Head of House	Marital Status	Gender	Occupation	Infirmity	Religion	Address
0	1	Kaur Drive	Douglas	Adams	77	Head	Widowed	Male	Retired Museum education officer	None	Christian	1 Kaur Drive
1	2	Kaur Drive	Hollie	West	79	Head	Single	Female	Retired Dealer	None	Christian	2 Kaur Drive
2	3	Kaur Drive	Joseph	Hall	55	Head	Single	Male	Fast food restaurant manager	None	Christian	3 Kaur Drive
3	4	Kaur Drive	Gregory	Cox	35	Head	Single	Male	Housing manager/officer	None	Not Religious	4 Kaur Drive
4	5	Kaur Drive	Lydia	Jones	42	Head	Single	Female	Unemployed	None	Christian	5 Kaur Drive
5	5	Kaur Drive	Francis	Jones	3	Son	NaN	Male	Child	None	Christian	5 Kaur Drive
6	6	Kaur Drive	Diana	Martin	80	Head	Widowed	Female	Retired Accounting technician	None	Christian	6 Kaur Drive
7	6	Kaur Drive	Ruth	Godfrey	12	Granddaughter	NaN	Female	Student	None	Christian	6 Kaur Drive
8	7	Kaur Drive	Glenn	Turner	76	Head	Widowed	Male	Retired Designer, textile	None	Not Religious	7 Kaur Drive
9	8	Kaur Drive	Suzanne	Jones	46	Head	Single	Female	Therapist, occupational	None	Not Religious	8 Kaur Drive

```

182 O'Connor Crossroad    22
5 Atkins Road              22
121 O'Connor Crossroad    22
134 O'Connor Crossroad    21
28 Crown Branch            21
..
4 Prince Parkways         1
5 Prince Parkways         1
7 Prince Parkways         1
9 Prince Parkways         1
14 Hull Street             1
Name: Address, Length: 3158, dtype: int64

```

```

array(['Morton', 'Singh-Morton', 'Marsden', 'Lewis', 'Thomas', 'Cook',
      'Thompson', 'Ward', 'Ali', 'Jenkins', 'Law', 'Smith', 'Hope',
      'Bell', 'Jones', 'Reynolds'], dtype=object)

```

→ Unique Surnames at
5 Atkins Road

There are 3,158 unique instances of physical addresses in the town, but this does not directly mean 3,158 households as some of the physical addresses appear up to 22 times, with multiple instances of unique surnames which suggests different households in an apartment complex or a block of flats.

The number of households was estimated by adding the number of individuals listed as heads of household to the number of lodgers. This assumes that every lodger makes up a single-person household, an assumption which was buttressed by the fact that all the lodgers in the census data were either single, divorced or widowed. With this metric, the total number of households was calculated to be 3,696. The occupancy level was derived by dividing the total population by the number of households and was found to be 2.7 individuals per household which is slightly above the UK average household size of 2.4 but does not necessarily indicate a housing crisis (ONS, 2021).

Recommendations

The strongest urgency of need was displayed by the high number of university students who need to commute to universities in neighboring towns and employed individuals who need to commute to their places of employment. As these commuters account for more than half of the entire population, I recommend building a train station to facilitate transportation and reduce the pressure on the existing road networks.

Observing the population pyramid and the age displot, the largest portion of the population consists of individuals in the prime of life and approaching retirement. This implies an unprecedented number of elderly people in the coming years, as these individuals progress into older age bands. I therefore also suggest that funds should be invested in old age care.

Bibliography

Harbour Family Law (2022) *Divorce Statistics*

Available Online: <https://www.harbourfamilylaw.co.uk/how-many-marriages-end-in-divorce-in-the-uk/>

[Accessed on 09 April 2023]

OECD (2023) *Employment rate by age group*

Available Online: <https://data.oecd.org/emp/employment-rate-by-age-group.htm>

[Accessed on 09 April 2023]

Office for National Statistics (2021) *Average Life Expectancy*

Available Online:

<https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/lifeexpectancies/bulletins/nationallifetablesunitedkingdom/2018to2020>

[Accessed on 09 April 2023]

Office for National Statistics (2022) *Childbearing Ages for Women*

Available Online:

<https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/conceptionandfertilityrates/bulletins/childbearingforwomenbornindifferentyearsenglandandwales/2020>

[Accessed on 09 April 2023]

Office for National Statistics (2021) *Families and Households in the UK*

Available Online:

<https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/families/bulletins/familiesandhouseholds/2020>

[Accessed on 09 April 2023]

Wells Burcombe Solicitors (2020) *Divorce Statistics*

Available Online: <https://www.wellsburcombe.co.uk/can-get-divorced-before-a-year-of-marriage>

[Accessed on 09 April 2023]

Wooley & Co Solicitors (2022) *Divorce Statistics*

Available Online: <https://www.family-lawfirm.co.uk/blog/how-long-does-a-divorce-take>

[Accessed on 09 April 2023]

World Health Organization (2023) *Crude Birth Rate*

Available Online: <https://www.who.int/data/gho/indicator-metadata-registry/imr-details/2978>

[Accessed on 09 April 2023]

World Health Organization (2023) *Crude Death Rate*

Available Online: <https://www.who.int/data/gho/indicator-metadata-registry/imr-details/3198>

[Accessed on 09 April 2023]