Introduction

The Census report provides an analysis of the population of a fictional town to make suggestions or recommendations for business and urban development on an empty plot of land. The first step of this process involves identifying and correcting errors such as empty or missing data, which is crucial for ensuring accurate and reliable analysis. By cleaning the data, town officials can make informed decisions based on a solid and trustworthy data

The following section in this report provides analyses and insights generated from the population data of the town.

Data Cleaning

The process of data cleaning commenced by reading the data and getting the necessary information about the data. Below shows that the census data has 11 columns and 8877 entries.

```
In [4]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8878 entries, 0 to 8877
Data columns (total 11 columns):
 #   Column                        Non-Null Count  Dtype
---  ------                        --------------  -----
 0   House Number                  8878 non-null   int64
 1   Street                        8878 non-null   object
 2   First Name                    8878 non-null   object
 3   Surname                       8878 non-null   object
 4   Age                           8878 non-null   float64
 5   Relationship to Head of House 8878 non-null   object
 6   Marital Status                6911 non-null   object
 7   Gender                        8878 non-null   object
 8   Occupation                    8878 non-null   object
 9   Infirmity                     8878 non-null   object
 10  Religion                      6857 non-null   object
dtypes: float64(1), int64(1), object(9)
memory usage: 763.1+ KB
```

Errors were identified in the following columns: Age, Relationship to Head of House, Marital Status, Gender, Occupation, Infirmity and Religion

The age column had float as its data type, it was converted to integer as only integer (single values) was considered as people's age in the population. There was no blank or missing data in the column. The Relationship to the Head of House column had four empty entries (3 adults and 1 minor). The 3 adults were assigned 'Head' based on the assumption that all adults can be a head of household, the minor was assigned 'None as there were no basis for assigning otherwise'.

The Marital Status column both had NaN and empty entries in it. The empty entry was assigned ‘Married’ based on his status as ‘Husband’ in the relationship to the head of house column. It was discovered that all the NaN entries in the column were people less than the age of 18 (minors), they were assigned ‘NA’ since they are all minors.

The Gender column had two empty entries; from close observation of other records, the two entries had the value ‘Son’ in the Relationship to the Head of House column, hence they assigned ‘Male’ as their Gender.

The occupation column had two blank entries (Lesley Green & Steven Norman) with Ages 71 and 46 respectively. The UK government has abolished the "default retirement age" and individuals can now continue working after reaching State Pension age (Gov.uk, n.d.). However, for the purpose of this analysis, all unemployed individuals above the age of 65 were assigned ‘Retired’. This same categorization was done for (Lesley Green) because there is no justifiable reason for assigning otherwise, given other records or data available for this individual. On the other hand, Steven Norman was assigned ‘Unemployed’ because he may be actually unemployed and had nothing to report as his occupation.

The Infirmity column had 8 empty entries, after close observation of these records, they were assigned ‘None’. These individuals may have decided to leave the field blank given that they had no medical condition.

The Religion column had the following unique values: Methodist', 'None', 'Catholic', 'Christian', nan, 'Muslim', 'Sikh', 'Jewish', 'Sith', 'Bahai', 'Private', 'Nope', 'Hindu. Close investigation revealed that two (2) records had ‘Sith’ and ‘Nope’ values. This was deemed to be an error as no such religions do exists. Furthermore, the column had 2021 NaN values, of which 51 were adults (18 and over) and 1970 minors (under 18). The adults are independent and could make decisions for themselves, hence they were assigned ‘None’. On the other hand, minors are recognized by law as dependents and could not make certain decisions for themselves. The religious affiliation of minors is often determined by the religion of their parents or legal guardians. Therefore, minors are typically assigned the religious belief of their parents. As a result, the Religion of the parents was assigned to their children using the "ffill" method as the data is arranged in such a way that the records of parents or legal guardians appear before their children's records in the data.

Population Demographics

Sequel to the completion cleaning, the census data will have the following columns.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8878 entries, 0 to 8877
Data columns (total 13 columns):
 #   Column                          Non-Null Count  Dtype
---  ------                          --------------  -----
 0   House Number                    8878 non-null   int64
 1   Street                          8878 non-null   object
 2   First Name                      8878 non-null   object
 3   Surname                         8878 non-null   object
 4   Age                             8878 non-null   int32
 5   Relationship to Head of House   8878 non-null   object
 6   Marital Status                  8878 non-null   object
 7   Gender                          8878 non-null   object
 8   Occupation                      8878 non-null   object
 9   Infirmity                       8878 non-null   object
 10  Religion                        8878 non-null   object
 11  Age_class                       8878 non-null   category
 12  Grouped Occupation              8878 non-null   object
dtypes: category(1), int32(1), int64(1), object(10)
memory usage: 807.1+ KB
```

```
In [187]: df.isnull().sum()

Out[187]: House Number                    0
          Street                          0
          First Name                      0
          Surname                         0
          Age                             0
          Relationship to Head of House   0
          Marital Status                  0
          Gender                          0
          Occupation                      0
          Infirmity                       0
          Religion                        0
          Age_class                       0
          Grouped Occupation              0
          dtype: int64
```

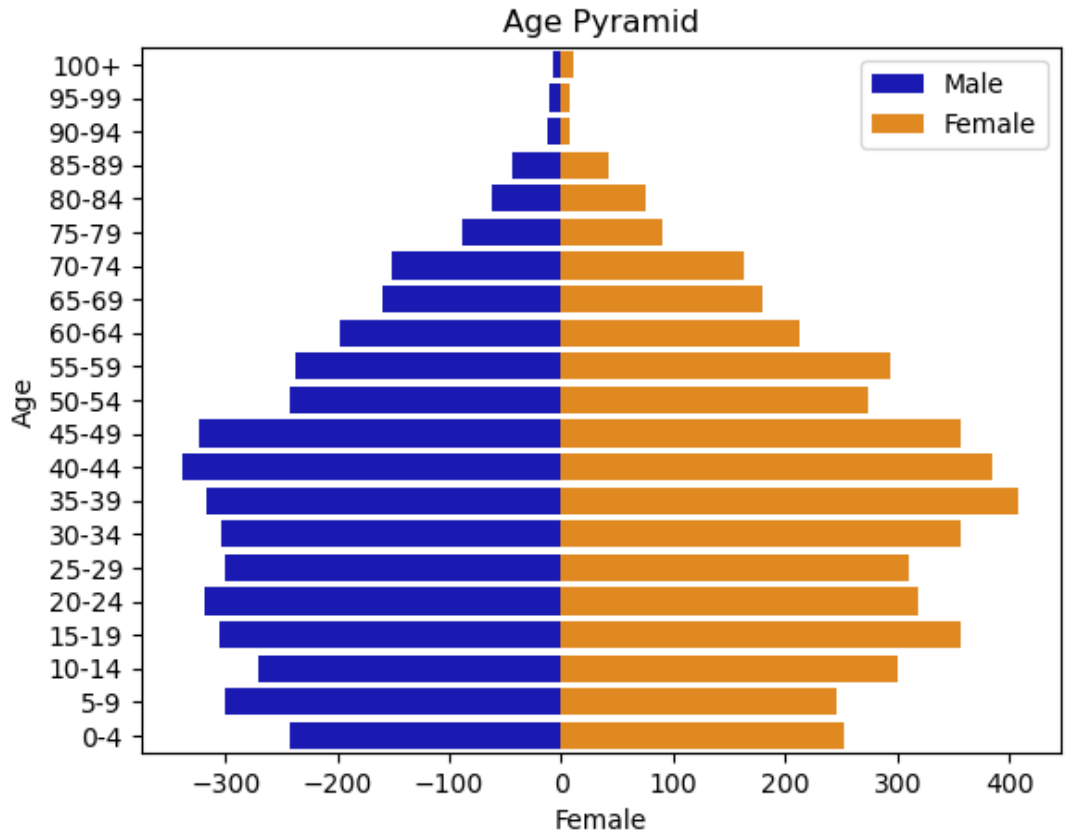
The following columns were added to support the further analysis of the data.

- Age_Class: The age group has been divided into intervals of 5 years to aid analysis in the data.
- Grouped Occupation: Occupation with similar attributes were grouped together.

The Age Pyramid

The population pyramid depicts a moderately high birthrate and number of school children (4 - 15). It has a high and active labour force (15-65). It is a growing population with a shrinking aged population. The pyramid further indicates that the majority of people in the population are employed, this could also be evidence of an affluent population. This is slightly different from the 2019 government statistics which shows a growing aged population

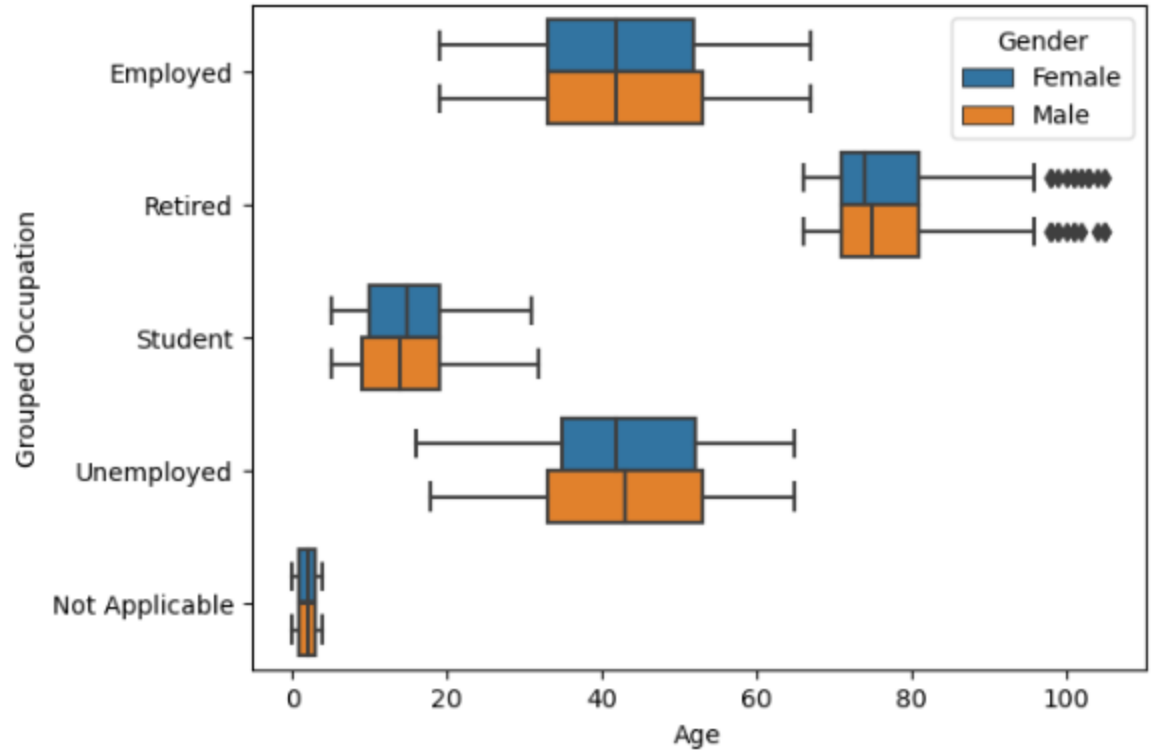
```
Text(0.5, 1.0, 'Age Pyramid')
```

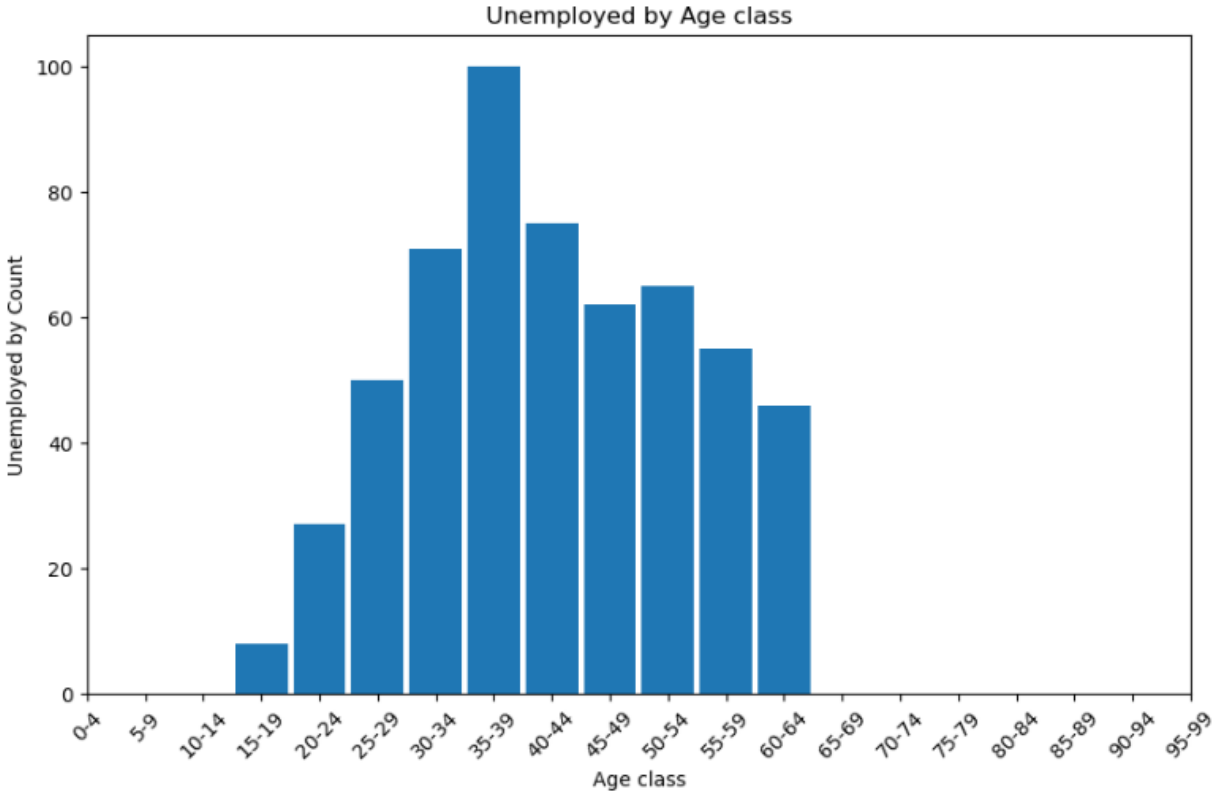


Employment Analysis

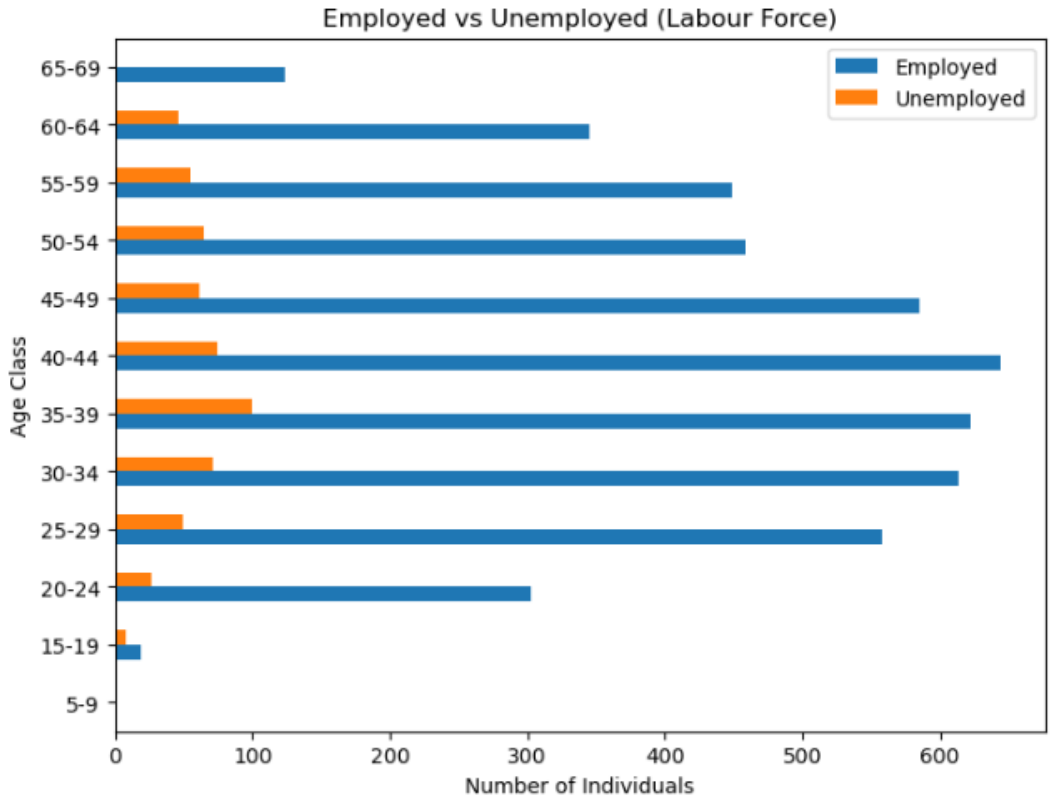
The scatterplot diagram below shows the plot for the different occupations conditioned with Gender. The plot shows high employment levels within the population, the majority falling with the age bracket (25-49). It shows unemployment levels is higher among female population than the male population.

```
<AxesSubplot:xlabel='Age', ylabel='Grouped Occupation'>
```





The records from the unemployment bar above chart shows that people who are above the age of 65 and are unemployed were inputted as 'Retired'. The chart shows a higher level of unemployment within the age bracket (35 -39).



The bar plot shows higher levels of employment within the population. This represents 53.2% of the total population as seen in the table below. The rate of employment (as a percentage of the labour force) stands at 89.4%. This is an indication of a wealthy population. The higher the employment levels, the higher the tax revenue for the government, and the overall national income. On the other hand, the unemployment level stands at 6.3% of the population and 10.5% of the working population.

Grouped Occupation

Value	Count	Frequency (%)
Employed	4720	53.2%
Student	2189	24.7%
Retired	916	10.3%
Unemployed	559	6.3%
Not Applicable	494	5.6%

Religion

Value	Count	Frequency (%)
None	3990	44.9%
Christian	2549	28.7%
Catholic	1313	14.8%
Methodist	733	8.3%
Muslim	149	1.7%
Sikh	72	0.8%
Jewish	60	0.7%
Private	5	0.1%

Marital Status

Value	Count	Frequency (%)
Single	3095	34.9%
Married	2552	28.7%
NA	1967	22.2%
Divorced	861	9.7%
Widowed	403	4.5%

Infirmity

Value	Count	Frequency (%)
None	8813	99.3%
Physical Disability	18	0.2%
Disabled	11	0.1%
Mental Disability	10	0.1%
Deaf	9	0.1%
Unknown Infection	9	0.1%
Blind	8	0.1%

Religion

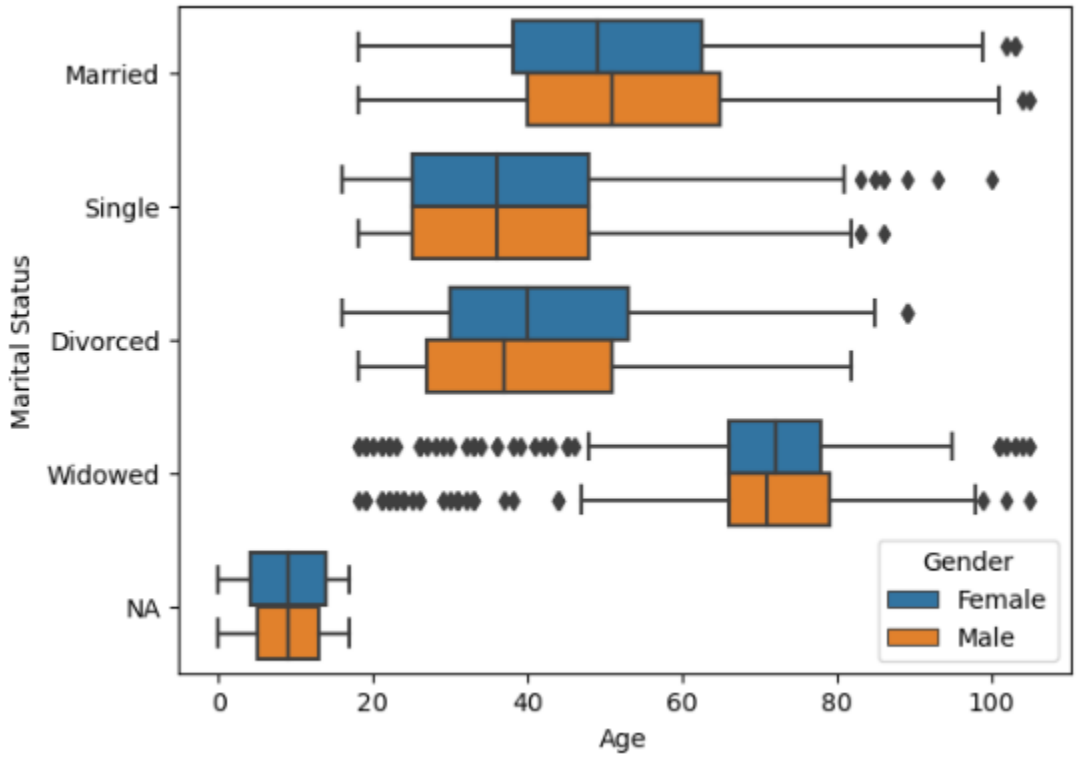
The descriptive statistics above shows that that the population is largely a non-religious population as close to half of the population (44.9%) have the value None as their religion. The next dominant and growing religion is Christianity. This figure could serve as a case for investing in developing a church building for Christians.

Infirmity

The infirmity column shows that 99.3% of people in the population have no infirmity or medical condition. This reflects the sound healthcare systems already existing in the town. As a result, there will be no need to invest in building a medical center.

Divorce and Marriage

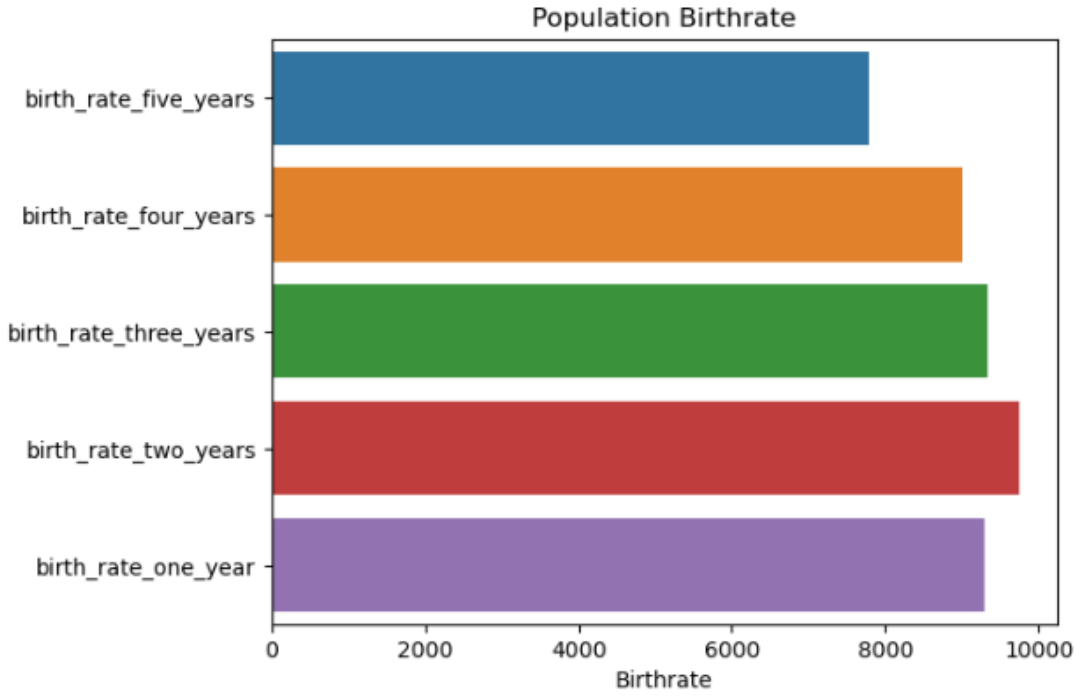
The percentage of married people in the population is 28.7% as seen the Marital status profile report. The bulk of these marriages occur within the middle-aged population. On the other hand, divorce stands at 9.7%, the box plot below shows that majority of divorces also occur within the middle ages. It further shows that there are more divorced females than males and that divorced men might move away from the town.



These figures is an indication of high divorce rates within the population and similar to the UK divorce rate in 2021 that put divorce rates in 9.3% for men and 9.4% for women per 1,000 of the population (Office for National Statistics, 2021).

Birth and Death Rate

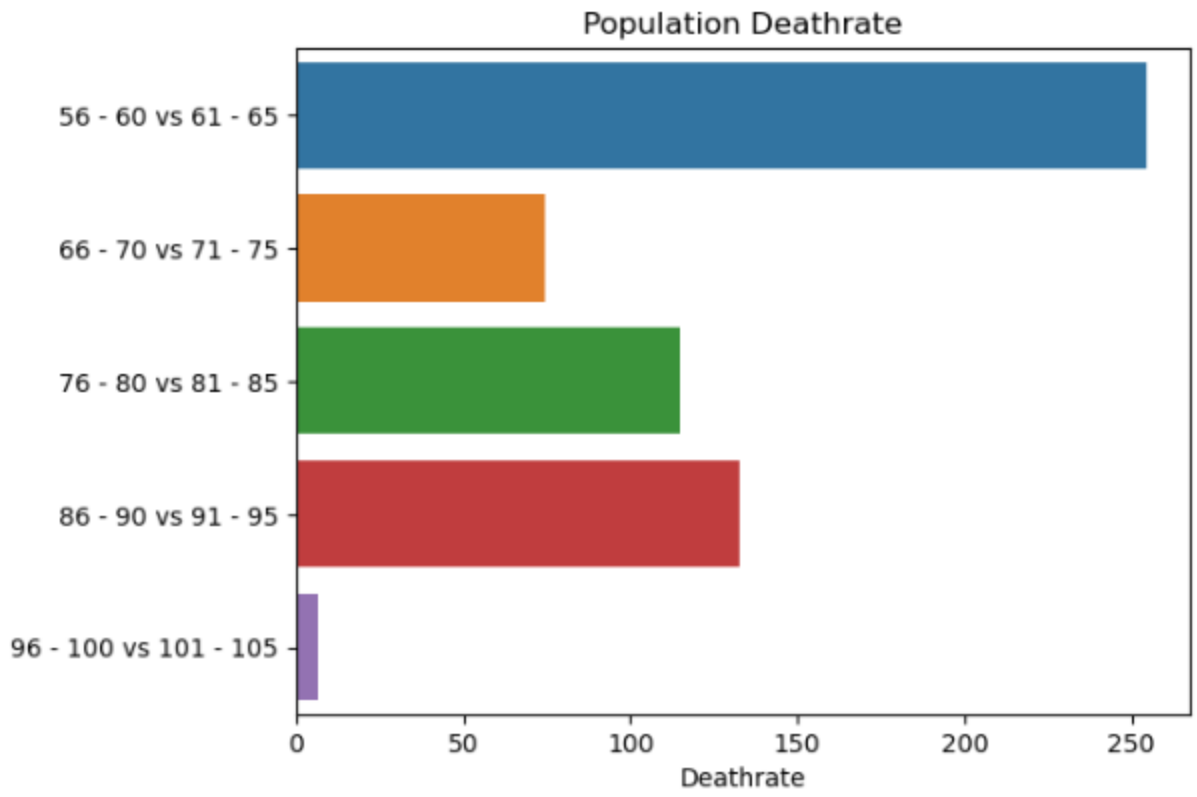
The population birthrate per hundred thousand has been increasing generally within the five-year period as seen below. It increased within the first four years and only declined marginally with the last or current year.



	Birthrate	Years
0	3790.613718	birth_rate_five_years
1	4365.436544	birth_rate_four_years
2	4596.665164	birth_rate_three_years
3	4802.513465	birth_rate_two_years
4	4676.258993	birth_rate_one_year

	Deathrate	Years
0	254.56	56 - 60 vs 61 - 65
1	74.34	66 - 70 vs 71 - 75
2	114.89	76 - 80 vs 81 - 85
3	132.91	86 - 90 vs 91 - 95
4	6.76	96 - 100 vs 101 - 105

The death rate per hundred thousand has an inconsistent rising trend. After the initial year, it increases for the next three years and falls sharply during the fifth. This reflects an overall rising death rate.



Immigration and Emigration

Some assumptions have been made about the migration from the population. The statistics for immigration consider lodgers and visitors who are not as divorced as immigrants. In contrast, emigration statistics are calculated by subtracting the number of male divorcees from the number of female divorcees. Based on the above, Immigration and emigration per hundred thousand stands out 2737 and 2106 respectively. This shows net migration is growing at 631 births per hundred thousand of the population as seen below.

```
# Net migration per 100,000

net_migration = ((immigration_count - emmigration_count) / Total_population) * 100000

net_migration
```

```
630.7726965532778
```

Overall, it was discovered that the population has been expanding significantly. It stood at 22,278 births per hundred thousand of the population.

The population growth rate is calculated by the mathematical equation below.

$$Population\ Growth\ rate = (total\_birth\_rate\ -\ total\_death\_rate)\ +\ net\_migration$$

Commuters

Commuters have been identified as people who live in the town and work in another town or city. The below set of occupations has been identified as potential commuters: University Students (including Ph.D. students), University professors (not retired), Other University staff (Librarians, Lab technicians), Sportsmen or professionals, Airline pilot, Military officers (i.e. Merchant navy). The town has been identified not to have any of these economic infrastructures to absorb these professionals. As such, they will need to move to other towns and cities for work.

These groups of people constitute about 14% of the labour force and 8.2% of the entire population.

Occupancy

The town has a total of 3384 houses and an average occupancy of 3 individuals per house, this is the average number of people expected to be living in any given house in the town given the present housing plan.

	House Number	Typical Count	Average Count
0	1	1	3
1	1	5	3
2	1	1	3
3	1	4	3
4	1	1	3
...	...	...	...
3379	232	1	3
3380	233	3	3
3381	234	1	3
3382	235	1	3
3383	236	1	3

The town has a total of 908 overcrowded houses which represents 27% of total houses in the town. On the Other hand, underoccupied houses stands at 1,938 representing 57.2% of the total houses. These figures show that more than half of the houses in the town are under-occupied even with the presence of lodgers and visitors.

Building further houses may not be needed at this time as these houses can be used to accommodate growing families, it could also be rented out to the rising number of immigrants (visitors, lodgers)

Recommendations

The census data for the town shows that there exists a high birthrate of 22,231 for the last 5 years and an overall population growth rate of 22,278 per hundred thousand births.

Further analysis of the population shows that 57% of the total number of houses are underoccupied. These unoccupied spaces could accommodate increasing family size and number of immigrants. Given this analysis, there exists no justification for building both high- and low-density houses. These funds could be channeled to fund other projects in the town.

Although the town is predominantly non-religious, there has been a noticeable increase in the number of Christians, specifically those who do not identify as Catholics. This suggests that constructing a church building that can accommodate all Christian denominations may be a justifiable investment.

Given that commuters constitute just 8.2% of the total population, it can be inferred that the majority of people who are employed live and work within the town. Hence, the need for building a train station may not be warranted.

Investing in general infrastructure may be necessary in light of the town's expanding population and the resulting strain on the existing infrastructure. The pressure on public facilities is likely to increase as more people are born and move into the town, making infrastructure improvements crucial for maintaining the quality of life for residents.

References

Gov.uk. (n.d.). State Pension age. Retrieved from https://www.gov.uk/state-pension-age [Accessed 11 April 2023].

Office for National Statistics. (2021). Divorces in England and Wales: Retrieved from https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/divorce/bulletins/divorcesinenglandandwales/2021[Accessed 10 April 2023].

Office for National Statistics (2022) Employment in the UK: Retrieved from https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/bulletins/employmentintheuk/july2022 [Accessed 19 April 2023].

Office for National Statistics. (2020). Childbearing for women born in different years, England and Wales: 2020. Retrieved from https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/conceptionandfertilityrates/bulletins/childbearingforwomenbornindifferentyearsenglandandwales/2020#:~:text=A%20woman's%20childbearing%20is%20assumed,not%20affect%20the%20overall%20patterns. [Accessed 27 April 2023].