CENSUS PROJECT REPORT

In order to establish plans, proposals, and suggestions for investments in future services and use cases for developments on an undeveloped plot of land, this report will be talking about the analyses of a modestly sized town between two larger cities given gathered data of the town. As a member of a local government team, I would also use this analysis to give the government precise demographic numbers that would help with planning, policy development, and money allocation.

The census data was first cleaned to address data inaccuracies and missing entries, as described in the first section of this study, before these suggestions could be made.

The report's subsequent sections will highlight significant analyses that were conducted with the express purpose of supporting the suggestions made. A summary of the town's population demographics is presented at the outset, and then a thorough study of the expected population growth, employment trends, commuters, and occupancy rates is provided.

Data Cleaning

To fix the data issues, census data was cleaned; a complete diary of all cleaning work is available in the associated Jupyter Notebook.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 11296 entries, 0 to 11295
Data columns (total 12 columns):
 #   Column                        Non-Null Count  Dtype 
---  ------                        --------------  ----- 
 0   Unnamed: 0                    11296 non-null  int64 
 1   House Number                  11296 non-null  object
 2   Street                        11296 non-null  object
 3   First Name                    11296 non-null  object
 4   Surname                       11296 non-null  object
 5   Age                           11296 non-null  object
 6   Relationship to Head of House 11296 non-null  object
 7   Marital Status                8408 non-null   object
 8   Gender                        11296 non-null  object
 9   Occupation                    11296 non-null  object
 10  Infirmity                     11296 non-null  object
 11  Religion                      8349 non-null   object
dtypes: int64(1), object(11)
memory usage: 1.0+ MB
```

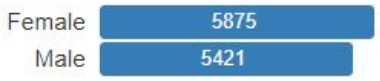
Info showing missing data in Marital Status and Religion columns

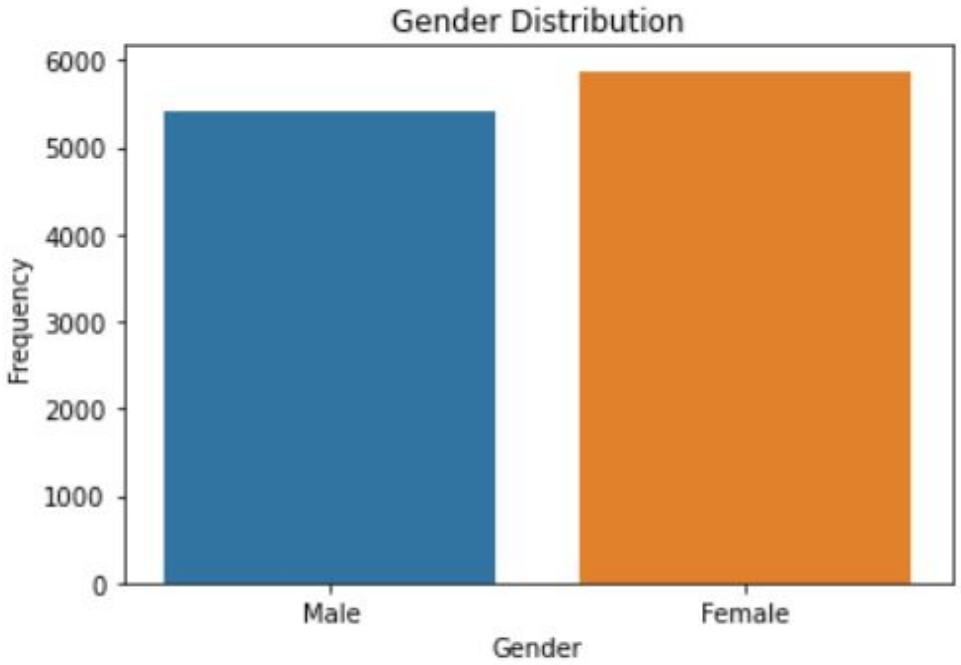
Necessary libraries needed for the data cleaning were imported (Pandas and seaborns). Checking the dataset for description, columns name, information of the data types in each column to make sure they are the right datatypes. I had some incorrect datatypes in House Number and Age columns which are

supposed to be in Integer instead of Objects. Also checking for the information of my dataset, I had some missing values in Marital Status and Religion.

Gender
Categorical

Distinct	2
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	88.4 KiB





Treating each column by checking the unique values so I can check for any irregularities. Starting with my Gender column, I had some misspells like ‘M’ should be Male and others which has been renamed. There were some empty spaces in my gender which I rectify using my domain knowledge of differentiating between male and female names. I can also deduce I have a total number of 5875 women and 5421 men in this population.

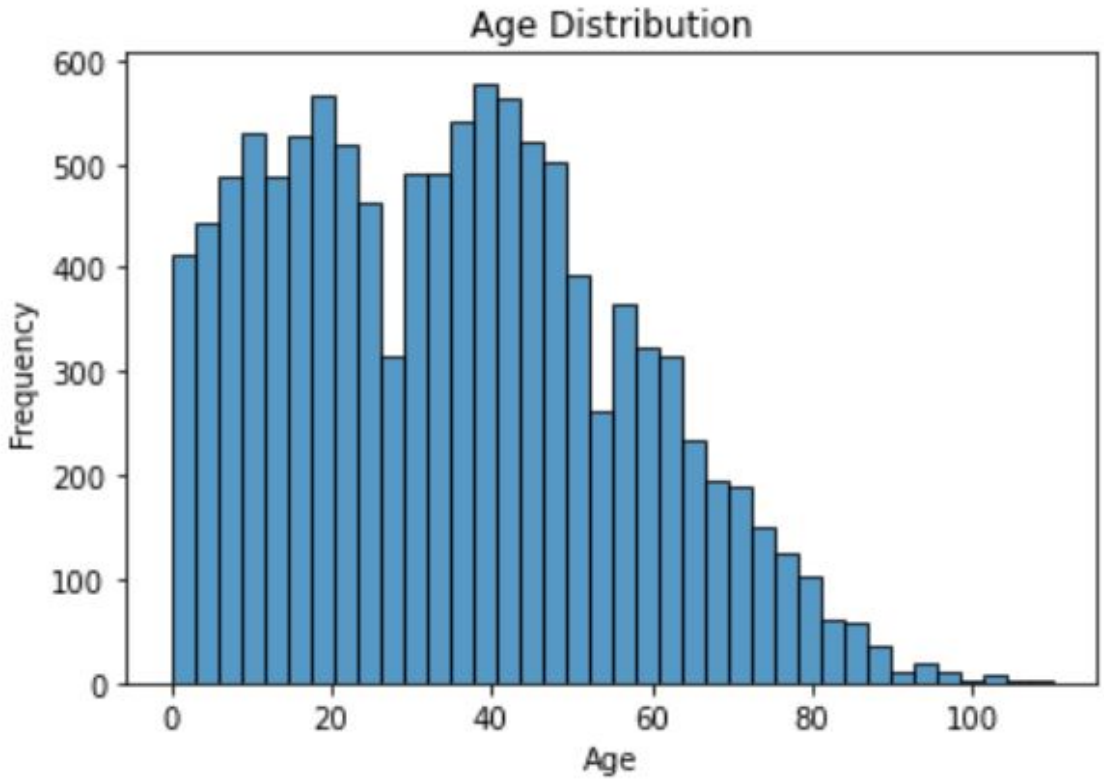


Chart showing age distribution in the population

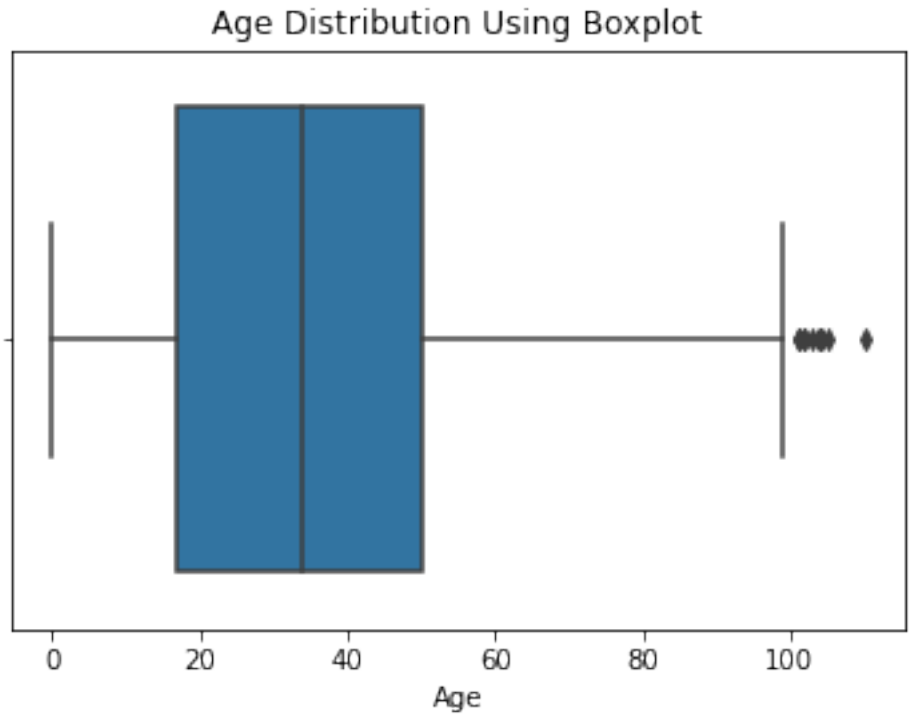
Treating my Age column, checking for the unique values; there are some error input like “Twenty” instead of “20”. More so Age column is in an object datatype which has been converted back to Integer.

Missing Values in the Marital Status are age under 18. Similar to adults, people under the age of 18 have a marital status of NaN. So All the Nan were replaced with Minor for this analyses.

There is also the same issue in House Number column just like the Age column where we have the word instead of number (One instead of 1) and the datatype is also object instead of Integer.

For the missing values in the religion column; after careful review, it shows there are Nan and the least age for missing values is 18 so we can simply replace the missing values with None since the people in the category are old enough to choose their own religion

Undecided and Housekeeper were converted to the religious category "None." The fact that there were only four entries for "housekeeper," which was known to be written as a "joke" entry in past censuses, led to the determination that this input was purposefully misleading (BBC, 2016). Undecided was changed to None because it's safe to presume you don't have a religion if you're unsure. When answering the survey for the dataset, the housekeeper was presumed to have been error or mistake as it only appears once in the dataset.



According to the boxplot distribution of the age, I can observe some outliers that were eliminated for the record of an age of 110 as shown above.

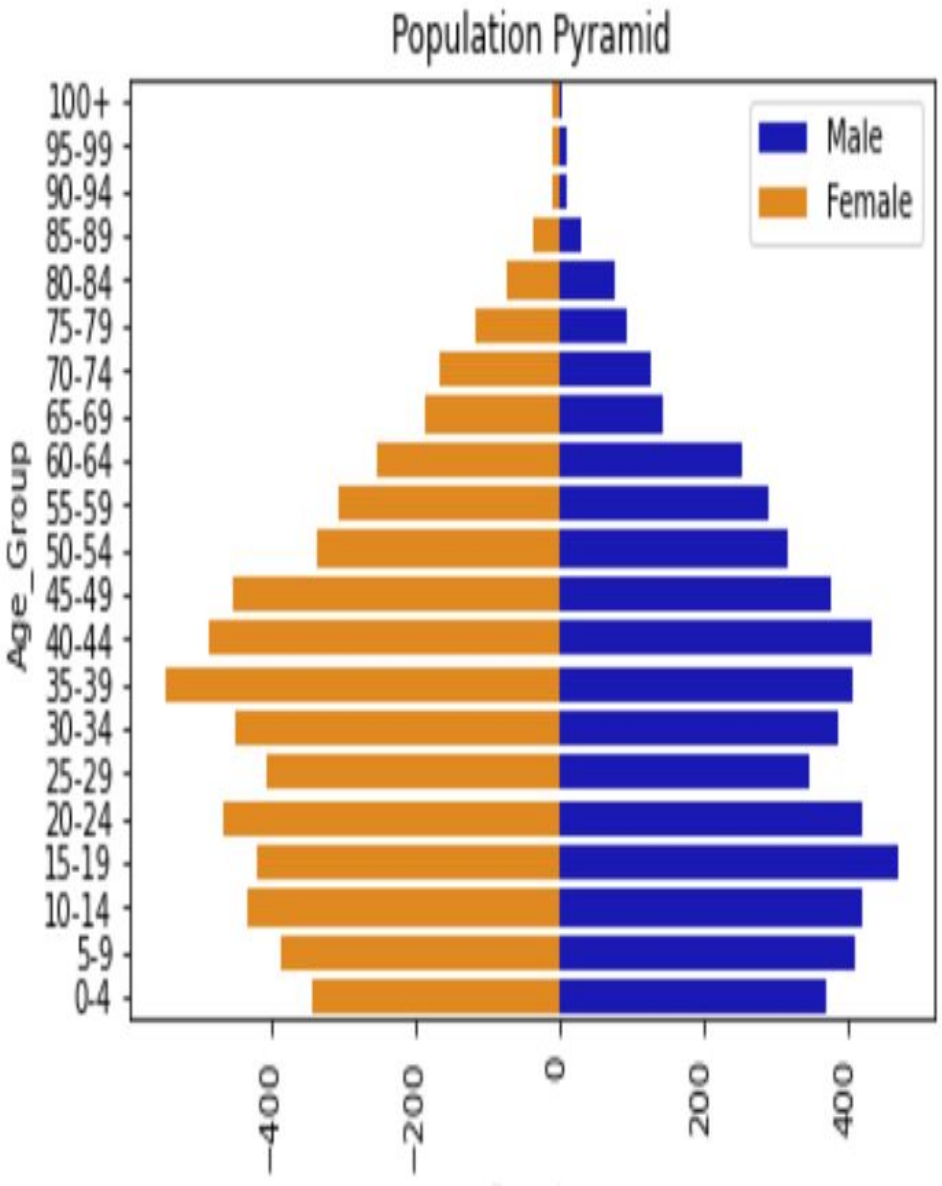
Population Demographics

Following data cleaning, the finalised census data will include the following characteristics below

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 11296 entries, 0 to 11295
Data columns (total 11 columns):
 #   Column                        Non-Null Count  Dtype
---  ------                        --------------  -----
 0   House Number                  11296 non-null  int32
 1   Street                        11296 non-null  object
 2   First Name                    11296 non-null  object
 3   Surname                       11296 non-null  object
 4   Age                           11296 non-null  int32
 5   Relationship to Head of House 11296 non-null  object
 6   Marital Status                11296 non-null  object
 7   Gender                        11296 non-null  object
 8   Occupation                    11296 non-null  object
 9   Infirmity                     11296 non-null  object
 10  Religion                      11296 non-null  object
dtypes: int32(2), object(9)
memory usage: 882.6+ KB
```

Info showing after cleaning the dataset

Demographics



Examining the population pyramid above, the structure of the population shows higher number of young people in the town. The town's population start shrinking from age 40 – 44. People are probably moving out of due to some reason we would be analyzing later in this report. Birth rate is moderately as we can see from aged 0-4. The population also tends to be shrinking for both male and female from aged 35 – 39; people are most likely to be moving out at certain. From the pyramid we must also considered building retirement home for old age as the population is shrinking. As there are more female than male in the population. There will be more young people in the future so the right investment in schools, old people retirement home and general infrastructure will lead to the development of this town

Detailed Analysis

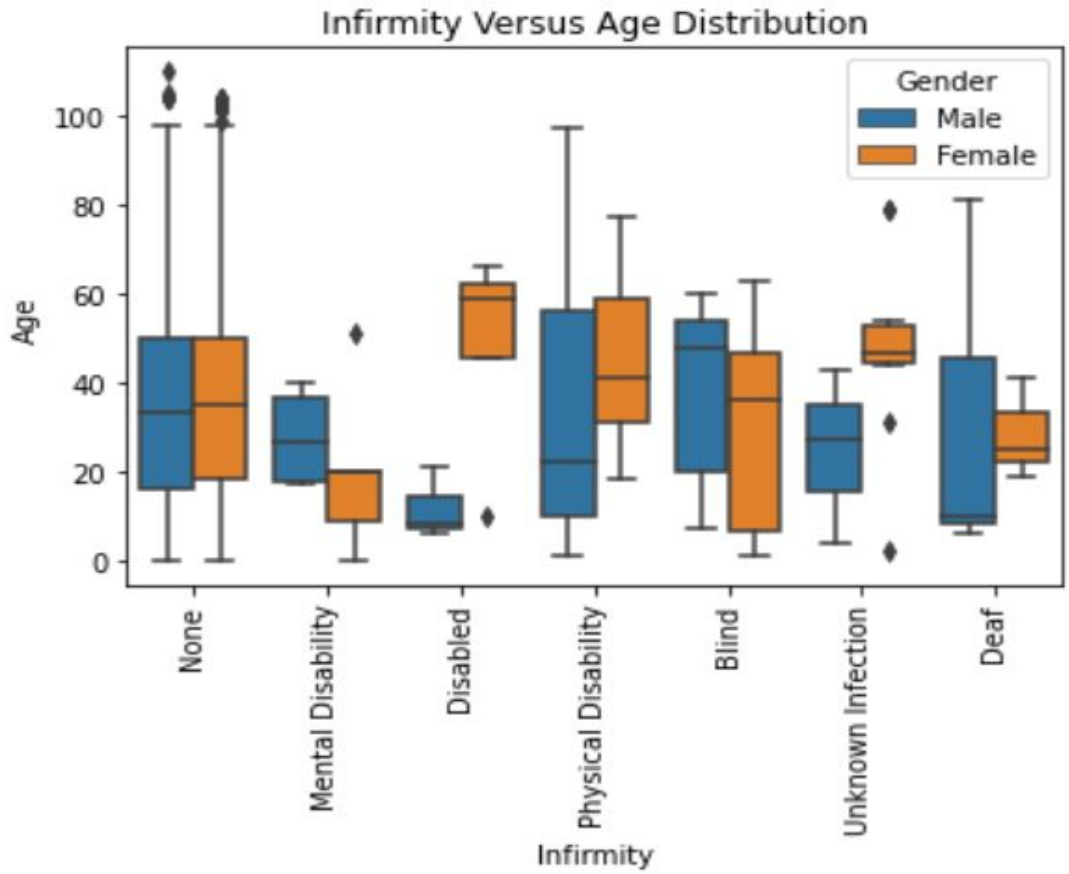
Infirmity

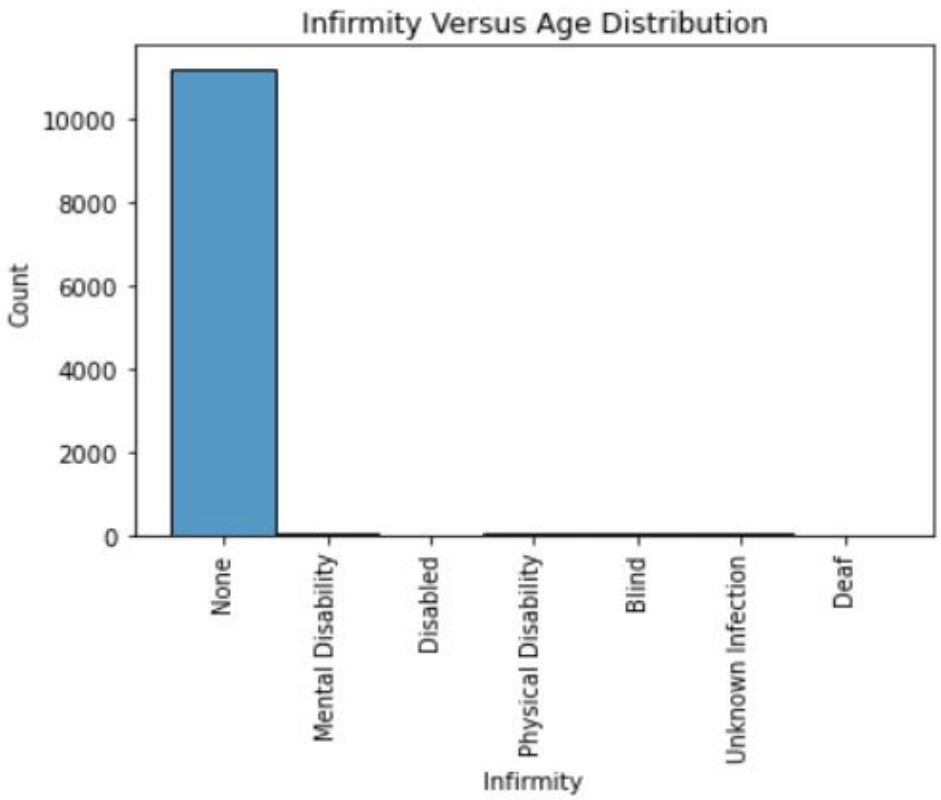
Infirmity
Categorical

Distinct	7
Distinct (%)	0.1%
Missing	0
Missing (%)	0.0%
Memory size	88.4 KiB

None	11221
Physical Disability	26
Unknown Infection	13
Blind	12
Mental Disability	11
Other values (2)	13

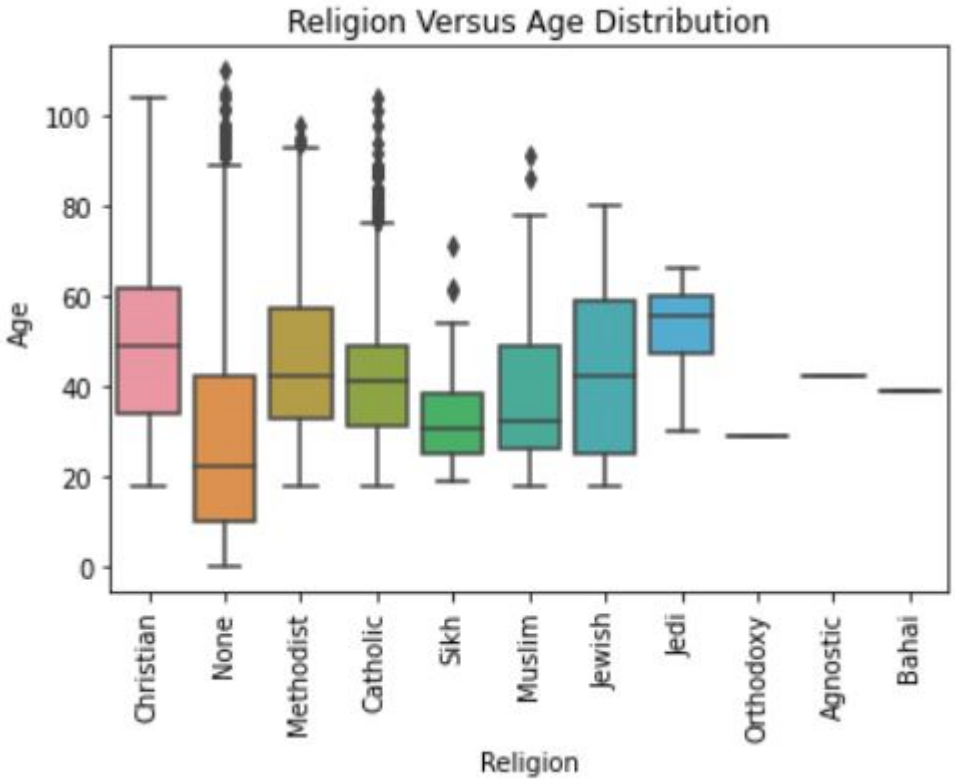
There will be no recommendations based on infirmity because it only made up a small portion of the population and was not regarded important enough to justify in-depth investigation. Regardless there's a room for improvement in the health sector if the town will be expanding. those with an infirmity in the town from the observation constituted less than 1% of the overall population of this town.





Religion

Analysing the religions in the town.



Some religions were expanding as shown in the chart above by the significantly younger age which are Sikhs and Muslims. These, too, made up a very minor portion of the population. Jewishness continues to be the most common religion, with both older and younger generations practising it. Muslims are increasing in this community; thus we might think about building a mosque to increase the population through this religion.

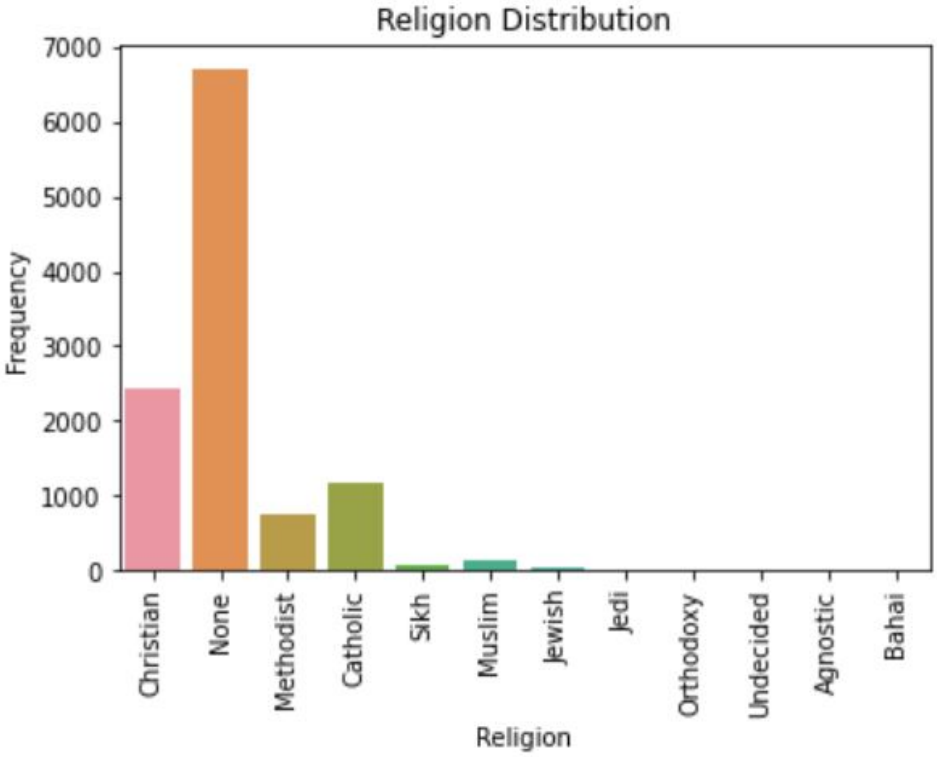
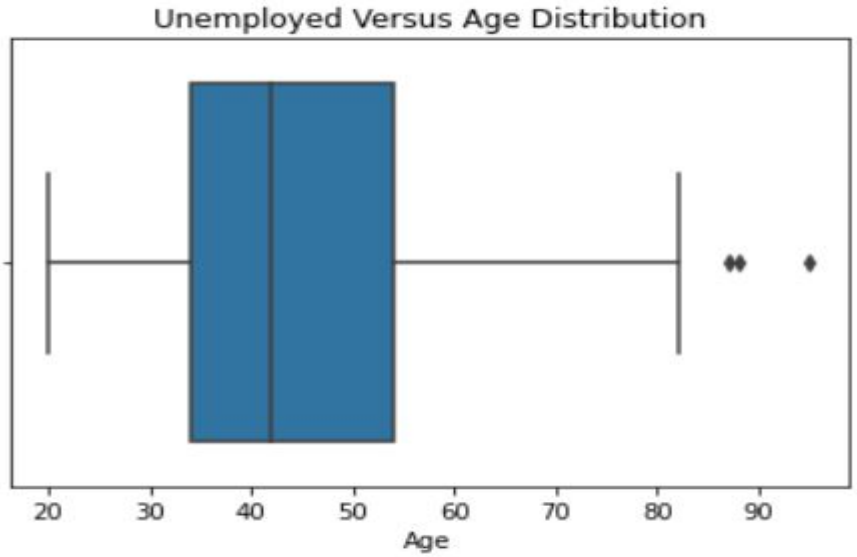
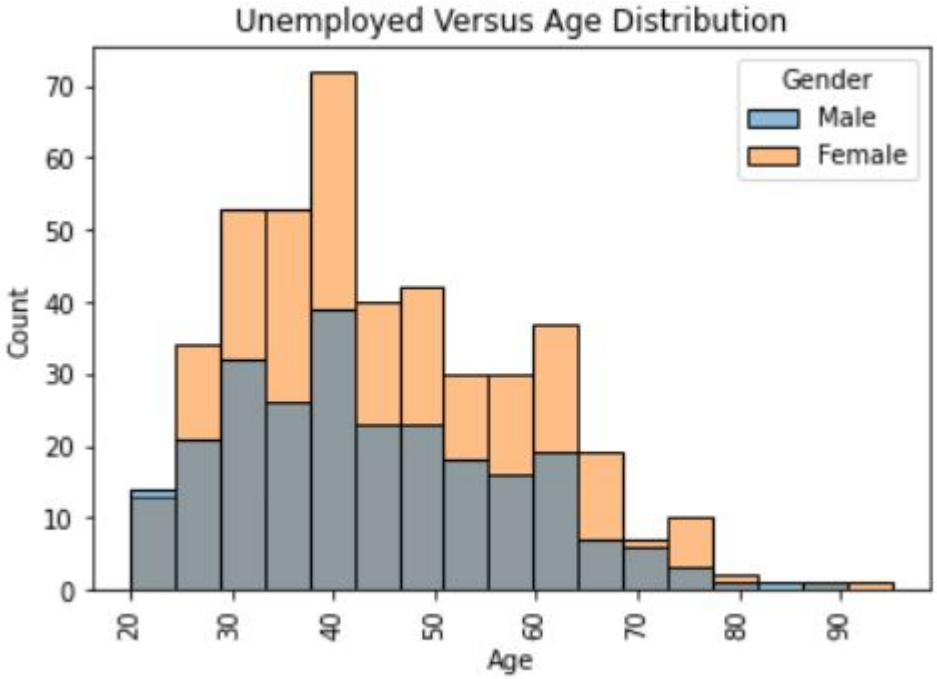


Chart showing we have more people with no religion in the population

Unemployment

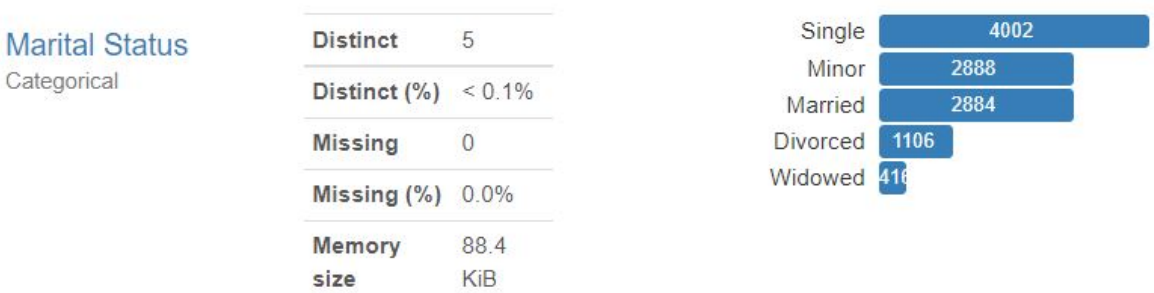


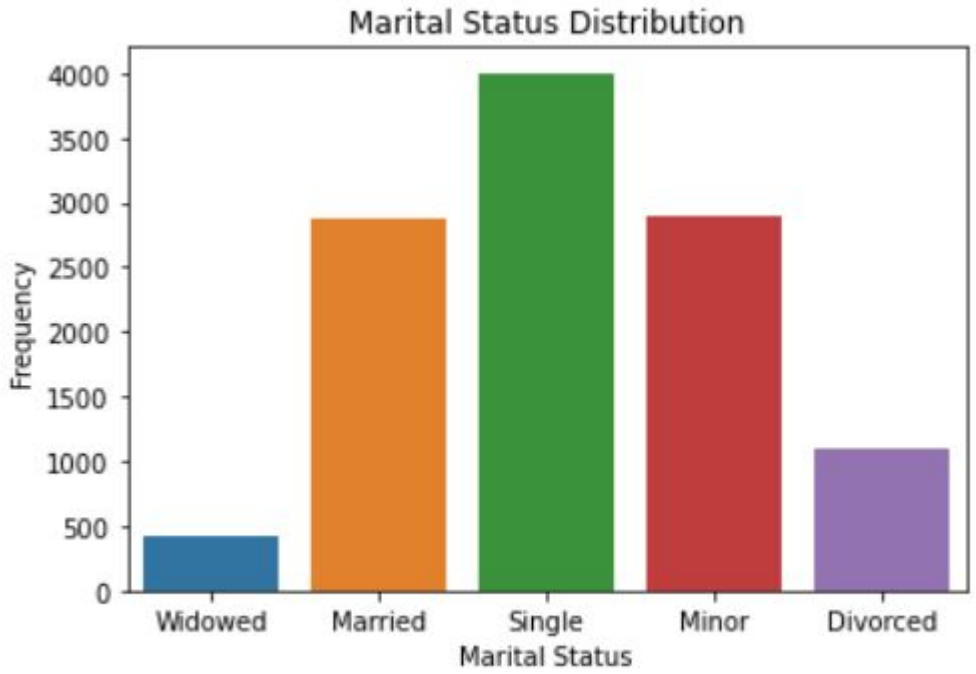
From the chart above. It is obvious that some people from the 34 – 54 have high rate of unemployment which will lead decrease in the population as the people are likely to move out of the town for seek job opportunities. Even if it is conceivable to experience unemployment after attaining retirement age, these cases shouldn't be taken into account in upcoming evaluations of unemployment because people over 65 are typically ineligible for employment.



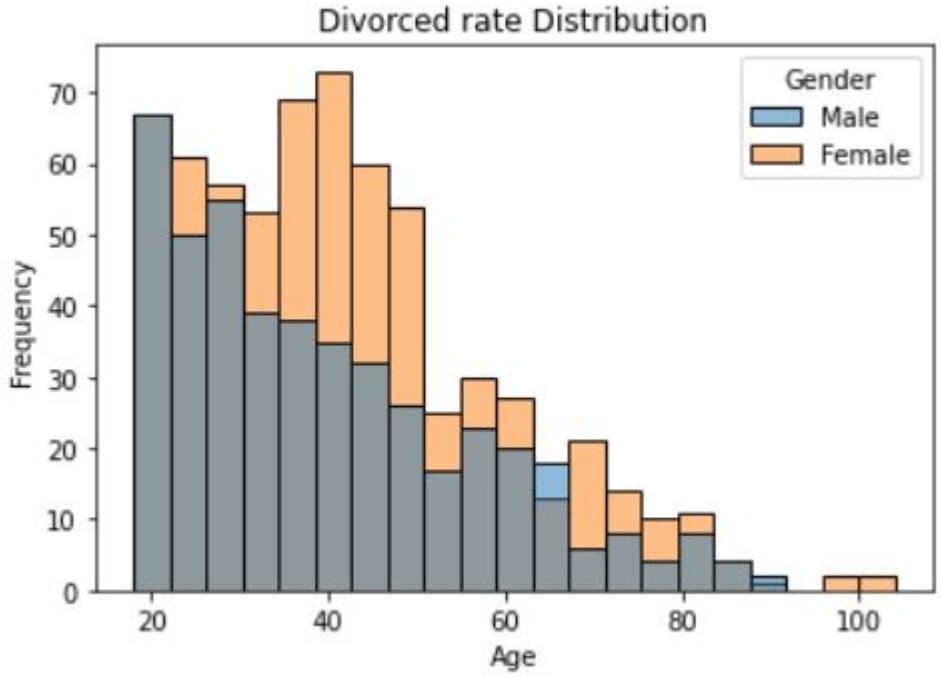
This charts shows more women are unemployed in this town. From the population pyramid, it shows we have more women in the town so we might have to be looking at empowering women to gain balance and expansion of the town. I can determine the unemployed rate has shown above in the Unemployed versus the age chart. It shows a lot of female are unemployed so I will be making some recommendations on that.

Divorce and Marriage





This chart shows there is a lot of single people in this population and we would be showing the relationship between married and divorce also from this dataset.



As was evident in the data cleaning section, divorce can happen at any age. There are more female divorcees than male, which suggests that male divorcees may depart the area, according to marital status data broken down by gender.

```
divorced_rate = len(df_divorced)/11296*100
divorced_rate
```

```
9.791076487252125
```

I can deduce from the dataset. We have 9.79% for the rate of divorce. Divorce rate is normal for the town

Marriage is calculated by counting the number of 'married' individuals and dividing by two. Using 1991–2004 data from the British Household Panel Survey (BHPS) analyses the effect of union dissolution on the occurrence of moves, changes of dwelling type, and the probability of moving out of owner-occupation. The main contributions of this paper are that it takes into account the rise in the occurrence of cohabitation, by analysing the dissolution of cohabiting and marital unions separately, and that it studies the effect of re-partnering on housing careers. Using logistic regression models clear evidence was found that the dissolutions of marriage and cohabitation result in different housing career outcomes. In particular, those who divorce experience a larger drop in housing quality than do those who split up from cohabitation. Starting a new relationship leads to more upward moves in the housing career compared to remaining divorced or split up.

Birth and Death Rate

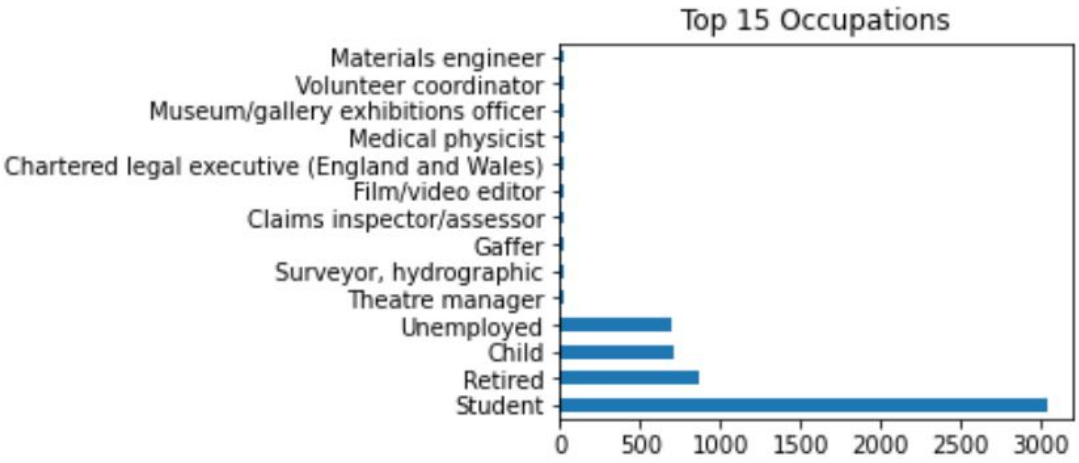
The birth rate for U.K. in 2021 was **11.377** births per 1000 people, a **0.49% decline** from 2020. The birth rate of this town is showing 1% higher.

```
birth_rate= (total_no_infants/11296)
birth_rate*1000
```

```
12.305240793201133
```

The birthrate is high for this town; this will lead to expansion. The birth rate and mortality rate are two key statistics in the study of demography. We must be proficient in these two numbers since they provide the foundation for understanding more intricate population concepts like growth rate, doubling time, and others. Fortunately, comprehending these two statistics is not too difficult: The birth rate is the number of babies born per 1,000 people, year. Similar to this, the death rate is the number of deaths but we might have to predict this as we don't enough information to use for death rate.

Occupation



University students as they make up the majority of emigrants and immigrants in the community, university students and students have merged together. According to the population pyramid chart above, the ages of 15 to 19 and 20 to 24 are both increasing, and the ages of 25 to 29 are decreasing as students are more likely to finish their education and move away owing to unemployment. The majority of commuters are in this group. Students should be seen as a constant in the expansion of the community because they depart the community after finishing their studies. This proves that the town will grow if a university is built there.

Lodgers and visitors are usually part of the calculate immigration figures. This is done to keep out divorcees who are staying elsewhere after divorcing their spouse and who wouldn't be considered newcomers to the area. However, emigration statistics are determined by comparing the number of male and female divorcees.

Commuters

Our dataset shows that there are many university students and students, which I have combined because they will all be travelling into and out of the town irrespective of the fact that there is no university in the town. This implies that the town will expand as more students are likely to remain there after school if there is a university in the town

Occupancy Rates

Showing the numbers of visitors and lodgers in the town below

```
df_visitor.count()
```

```
House Number                    28
Street                          28
First Name                      28
Surname                         28
Age                             28
Relationship to Head of House   28
Marital Status                  28
Gender                          28
Occupation                      28
Infirmity                       28
Religion                        28
New Age                         28
dtype: int64
```

```
df_lodger.count()
```

```
House Number                    577
Street                          577
First Name                      577
Surname                         577
Age                             577
Relationship to Head of House   577
Marital Status                  577
Gender                          577
Occupation                      577
Infirmity                       577
Religion                        577
New Age                         577
dtype: int64
```

```
#the number of members per household
df[['House Number', 'Street', 'Surname']].value_counts()
```

```
House Number  Street            Surname
59            Beverley Drive    Hancock     22
38            Daffodiltug Road  Murray      22
49            Holyrood Lane     Ryan        21
39            Holyrood Lane     Stevens     20
91            Beverley Drive    Richards    14
                                            ..
23            Wharf Lane        Ward         1
4             Hall Spurs        Lloyd        1
23            Wilson Lane       Green        1
                                Parry        1
24            Moss Road         Wilson       1
Length: 5678, dtype: int64
```

Families with high incomes living together and students could be the cause of extremely high occupancy, albeit these could also be apartment buildings. According to the example given above, some homes might be overcrowded with 22 occupants, while the other situation might involve an affluent family. However, there are reportedly 28 guests and 577 lodgers who live with families. There are three possible explanations for this: families are sharing rooms in their overcrowded homes to make way for lodgers; families are unable to downsize.

Recommendations

The town has a lot of students, who are the main commuters, therefore there could be a lot of commuters there. I suggest investing in transportation by establishing a railway station that connects the town to the nearby cities.

Increase education funding as there is evidence of an increase in the number of children who are old enough to attend school (new births). Expansion will result from university development investments since more students will choose to stay in the community after completing their studies.

Broader infrastructure Services (such as waste collection, road upkeep, etc.) will demand more expenditure if the town is growing. Mainly general infrastructure.

The town will have a strong influx of lodgers and visitors, which will increase the number of commuters. Additionally, I advise considering making an investment in low-density housing.

It demonstrates that the town's residents enjoy good health, but if the municipality is to expand, we must also make investments in the healthcare industry.

Also If funding for transportation, health care, and education is taken into account, funding for training and population employability must also be taken into account.

REFERENCES

U.K. Birth Rate 1950-2022 | MacroTrends

Feijten, Peteke, and Maarten Van Ham. "The impact of splitting up and divorce on housing careers in the UK." *Housing Studies* 25.4 (2010): 483-507.

Feijten, P., & Van Ham, M. (2010). The impact of splitting up and divorce on housing careers in the UK. *Housing Studies*, *25*(4), 483-507.

Feijten, P. and Van Ham, M., 2010. The impact of splitting up and divorce on housing careers in the UK. *Housing Studies*, *25*(4), pp.483-507.

Feijten P, Van Ham M. The impact of splitting up and divorce on housing careers in the UK. Housing Studies. 2010 Jul 1;25(4):483-507.

https://www.ons.gov.uk › divorce › bulletins › 2021