INTRODUCTION

This is the comprehensive final report of the mock census of an imaginary modestly sized town.

The mock census contains randomly generated data using the Faker package in python. The data was generated in a similar manner to the 1881 census.

The aim of the project is to clean, analyse and make recommendations as some officials of a local government pertaining to unoccupied plot of land and to decide on investing in either Employment training, old age care, increased funding for schooling or general infrastructure.

DATA CLEANING PROCESS

A summary of the census dataset 'censuso3.csv' is shown in figure 1.

```
# Census data information showing data columns,Non-Null values Count and Data types.
df.info()
```

```
<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  int64 
 1   Street                        10083 non-null  object
 2   First Name                    10083 non-null  object
 3   Surname                       10083 non-null  object
 4   Age                           10083 non-null  object
 5   Relationship to Head of House 9424 non-null   object
 6   Marital Status                7526 non-null   object
 7   Gender                        10083 non-null  object
 8   Occupation                    10083 non-null  object
 9   Infirmity                     88 non-null     object
 10  Religion                      4100 non-null   object
dtypes: int64(1), object(10)
memory usage: 866.6+ KB
```

Figure 1: Summary of Census Data

The dataset contains Ten thousand and eighty-three (10,083) entries on eleven (11) columns. The columns for Relationship to Head of Household and Marital Status have some missing values. As for the columns for Infirmity and Religion, we assume that the computer has attributed the answer of 'None' to the census question as missing values.

Children below the legal age of marriage were changed to 'Not of Marriage age' in the marital status column, while 'No Religion' was assigned to anyone below the age of eighteen (18) who did not fill the religion column.

Misspelling in the religion column was corrected, while missing first name were labelled as 'undeclared'.

'Christian' in the Religion column was re-labelled as the church of England(Church of England - Wikipedia)

Th missing value in the Age column was replaced with the median age, while corrections were made for datatype and whitespaces.

A self-generated function **house info ()** which returns information on specific entries was used especially for the relationship to the head of house columns.

In the relationship to head of house column Lodger and Visitor where merged since they were the only non-family relation in that column.

The link below is the Jupiter notebook file showing the step-by-step process of cleaning the data. This file has also been attached as one of two Jupiter notebook files submitted for this report.

http://localhost:8888/notebooks/UNDERSTANDING%20DATA%20SCIENCE%20PROJECT_DATA%20_CLEANING.ipynb

The summary of the cleaned census data 'census03_clean' is shown in figure 2.

```
# Displays the data type, non null count and number of data entry
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10083 entries, 0 to 10082
Data columns (total 12 columns):
 #   Column                          Non-Null Count  Dtype
---  ------                          --------------  -----
 0   Unnamed: 0                      10083 non-null  int64
 1   House Number                    10083 non-null  int64
 2   Street                          10083 non-null  object
 3   First Name                      10083 non-null  object
 4   Surname                         10083 non-null  object
 5   Age                             10083 non-null  int64
 6   Relationship to Head of House   10083 non-null  object
 7   Marital Status                  10083 non-null  object
 8   Gender                          10083 non-null  object
 9   Occupation                      10083 non-null  object
 10  Infirmity                       10083 non-null  object
 11  Religion                        10083 non-null  object
dtypes: int64(3), object(9)
memory usage: 945.4+ KB
```

Figure 2:Summary of clean census data

OVERVIEW

The census data of this town shows that it has a population of about Ten thousand and eighty-three (10,083) people.

There are five thousand two hundred and sixty-seven males representing fifty-two percent (52) of the population and four thousand eight hundred and sixteen females representing forty-eight (48) percent of the population as shown in figure 3.

About Eleven percent of the residents of this town are aged 65 years and over and are mostly retired from working.

Sixty-six percent (66) of the population are within the working age of 18-64, While there is a high proportion of young people aged 20-24 largely due the university in the next town. They account for eight (8) about percent of the town's population as shown in figure 4.

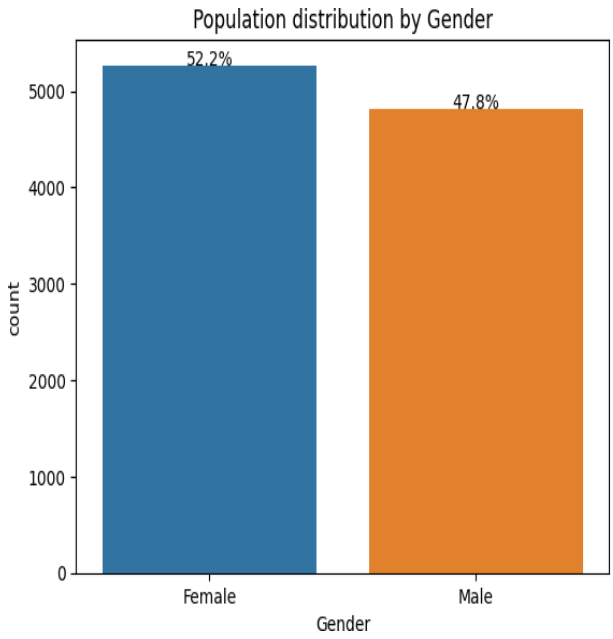


Figure 3: Population distribution by Gender

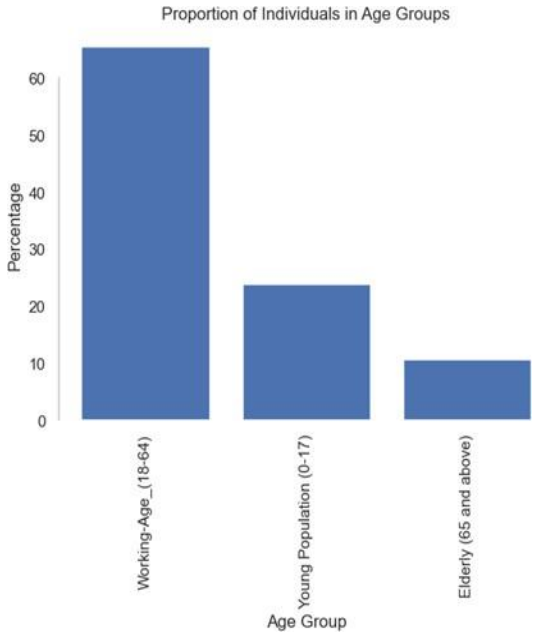


Figure 4: Proportion of Population by age groups.

AGE PYRAMID

The age pyramid is shown in figure 5, there are twenty-three (23) age groups each with an age range of five (5). The minimum age band was 0-4, while the maximum was 110-114.

From the pyramid we observe a steady increase in the population from age range 0-4 up to age range 15-19. This could be due of increased birth rates in those years or families with early school age kids moving into the town.

There is a slight decrease in the population in age group 20-24 relative to age group 15-19. This could be because of parent's migration or some of the newly admitted university students deciding to stay on campus.

There is further reduction in the population between ages 20-24 and 25-29. This might be attributed to university graduates deciding to move from the town after graduation or young adults leaving town for new job opportunities.

There is a higher representation of women in the reproductive age group 15-49 which follows a common demographic pattern.

Age groups between ages 35 and 49 constitute the largest proportion of the population. This indicates a significant middle age population.

Age 65 is the retirement age. However, there is high proportion of the population in the 55-=59 and 60-64 age group that will become retired in the next ten years.

There is a consistent decline in population from age 60-64 to 1100-114.This may be due higher mortality among elderly population, migration etc.

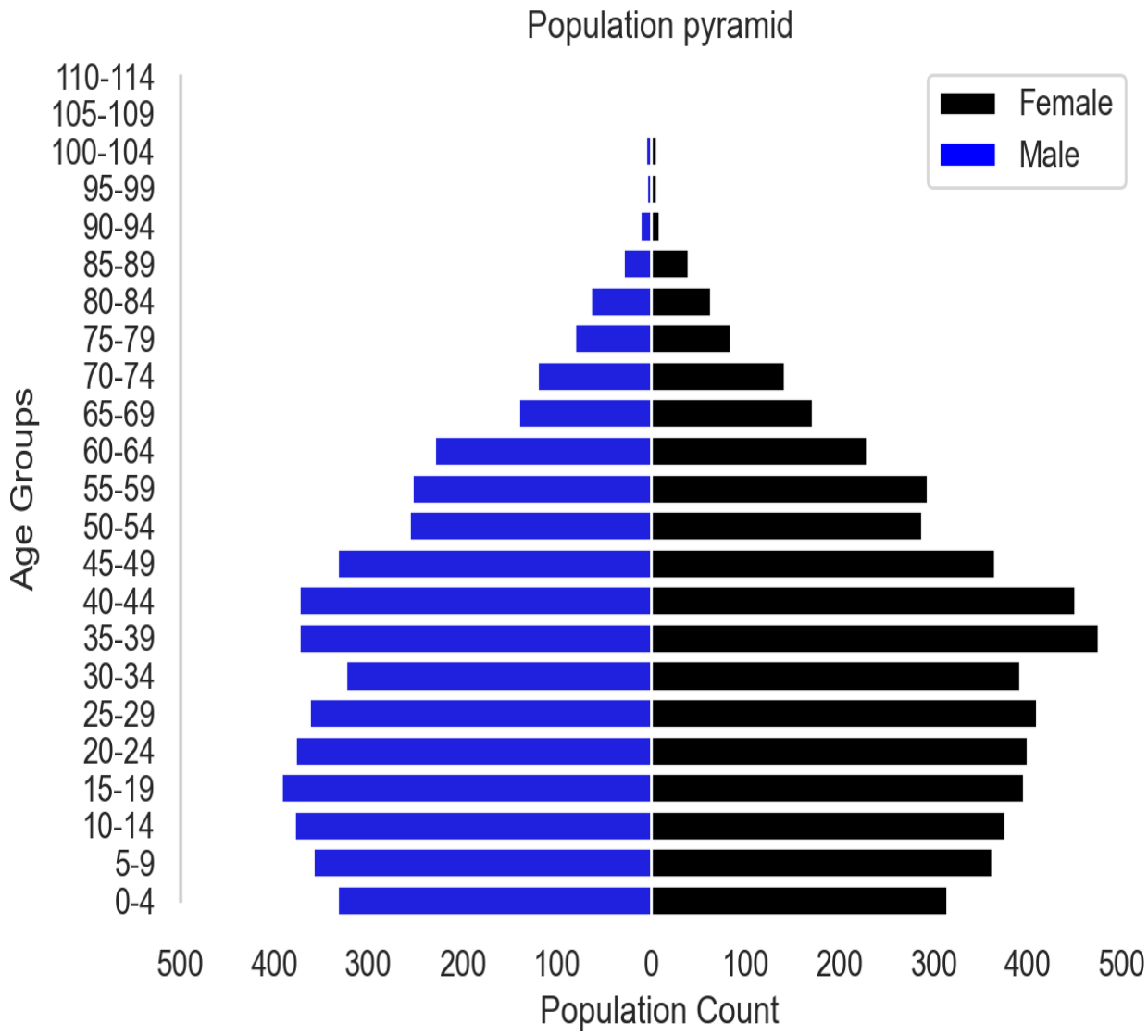


Figure 5: Age Pyramid

MARITAL STATUS

From figure 6, thirty-Six (36) percent of the population are single, while twenty-five (25) percent are married and a further twenty -five (25) percent are not of marriage age. The divorce rate is 9.64% while the marriage rate is 25.30%.

The crude divorce-rate for a particular year is calculated by dividing the number of divorces occurring within a population over the year, by the average or mid-year population for that year, expressed times 1,000.

Divorce rates in the 2021 census were 9.3 for men and 9.4 for women per 1,000 of the married population (including both opposite-sex and same-sex couples).(ons.gov.uk)

In this study the percentage of divorced female is 5.7 percent compared to 3.9 percent for men. We assume that either the female divorcees migrated to the town, or the male divorcee left the town. In contrast, there is no significant difference in the marriage rate amongst male and female population. There `are more single females in the population than single males. This is shown in figure 7.

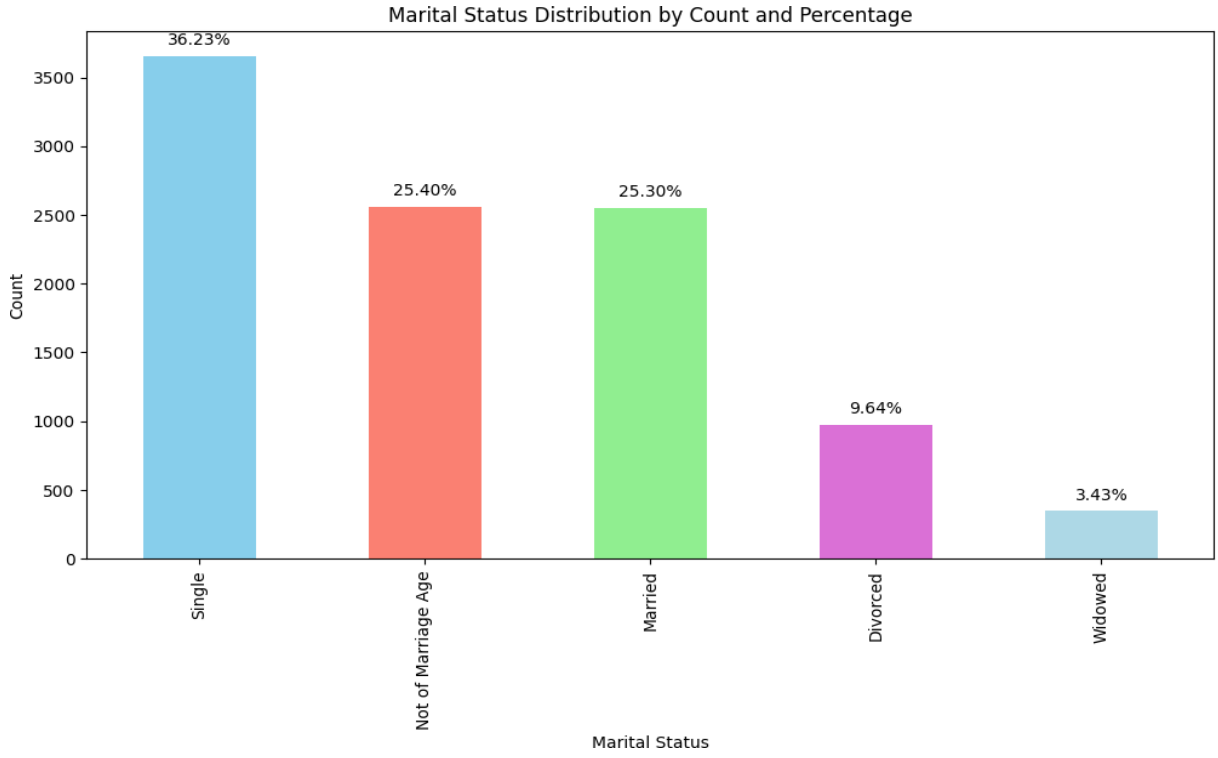


Figure 6:Distribution of Marital Status

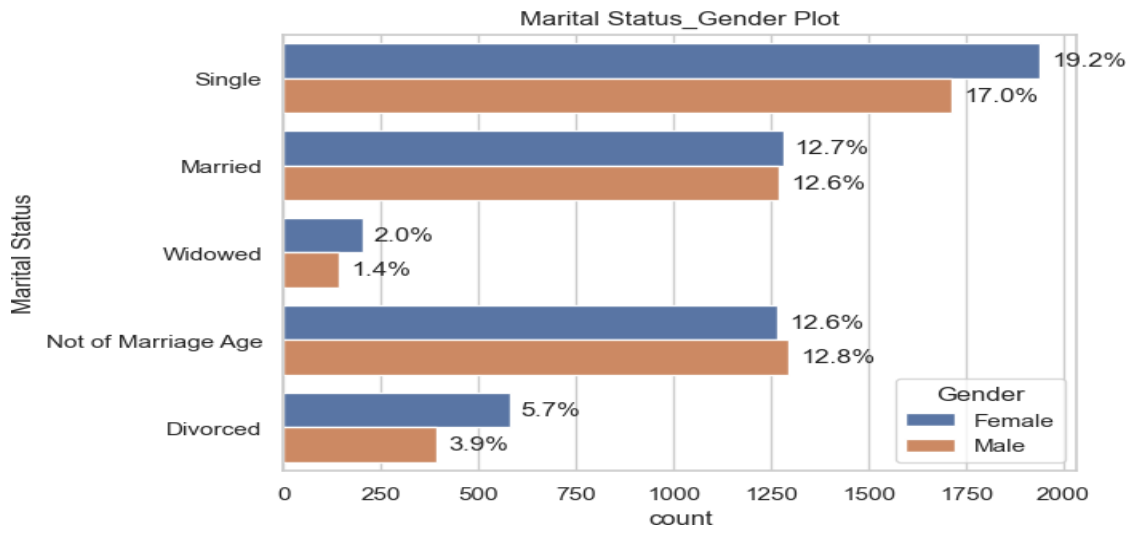


Figure 7:Marital Status by Gender

OCCUPTION

Fifty-two (52) percent of the population were employed, while six (6) percent are unemployed. The unemployment rate is 10 percent, while the employment rate is ninety percent. A closer look at the unemployment trend shows that the highest proportion of unemployed are in the age range 35 -44, followed by the age range 25 – 34, this might be predominately fresh graduates.

About 8 percent of the population are retired, while about fourteen (14) percent of the retirees above sixty-five (65) years old. does some form of paid work after attaining the retirement age. This shown in figure 8.

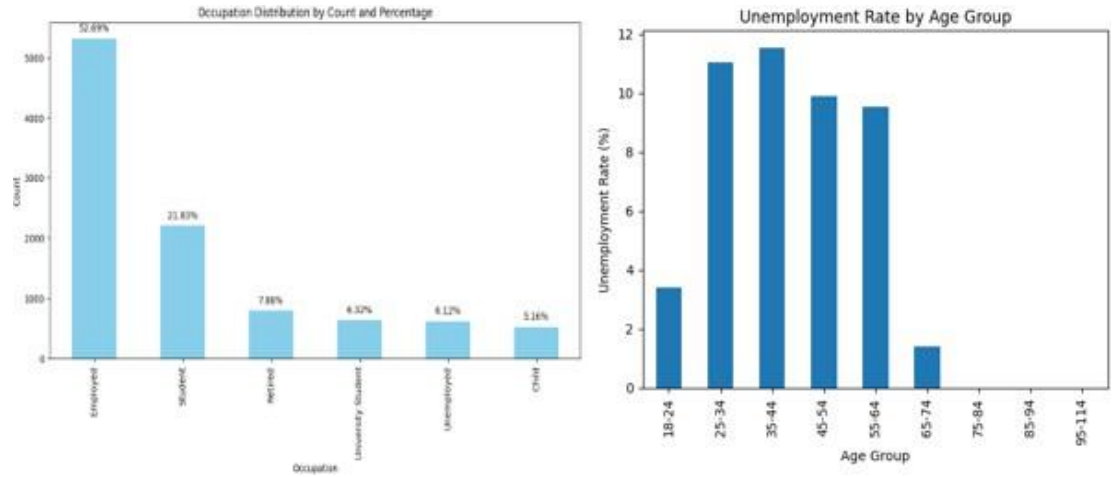


Figure 8: Occupation and Unemployment

RELIGION

For this study, the church of England or Anglicans were assumed to be Christians while other Christian denominations remained as captured in the census.

“No religion” was the most common response in this census count. About fifty-nine (59) percent of respondents had no religious affiliation.

The church of England or ‘Christian’ adherents accounted for about twenty-two (22) percent of the respondents.

Figure 9 below shows that respondents who had no religious affiliation had the lowest mean age compared to ‘Christian’ and catholic respondents.

The 2021 census of the of the United Kingdom suggests that Christianity is no longer the majority religion in England and Wales (guardian, 2022). This they attributed to an ageing population.

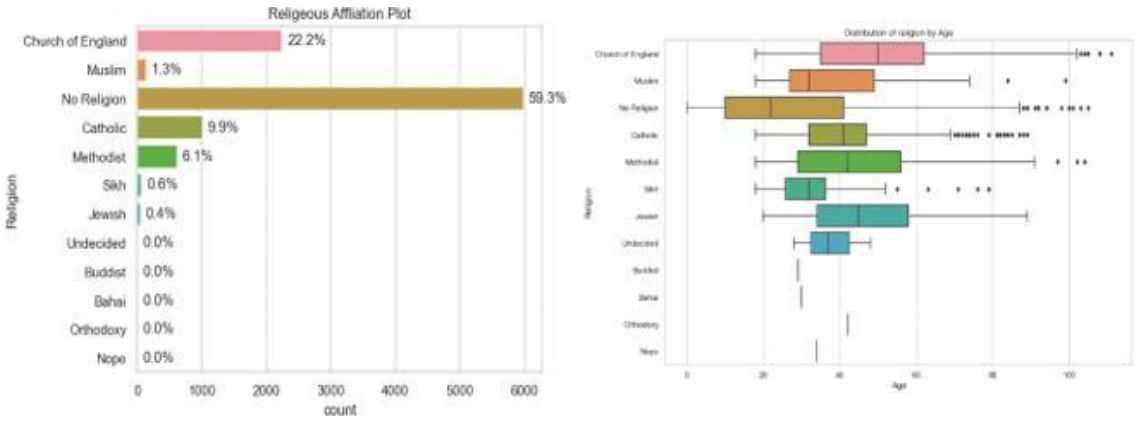
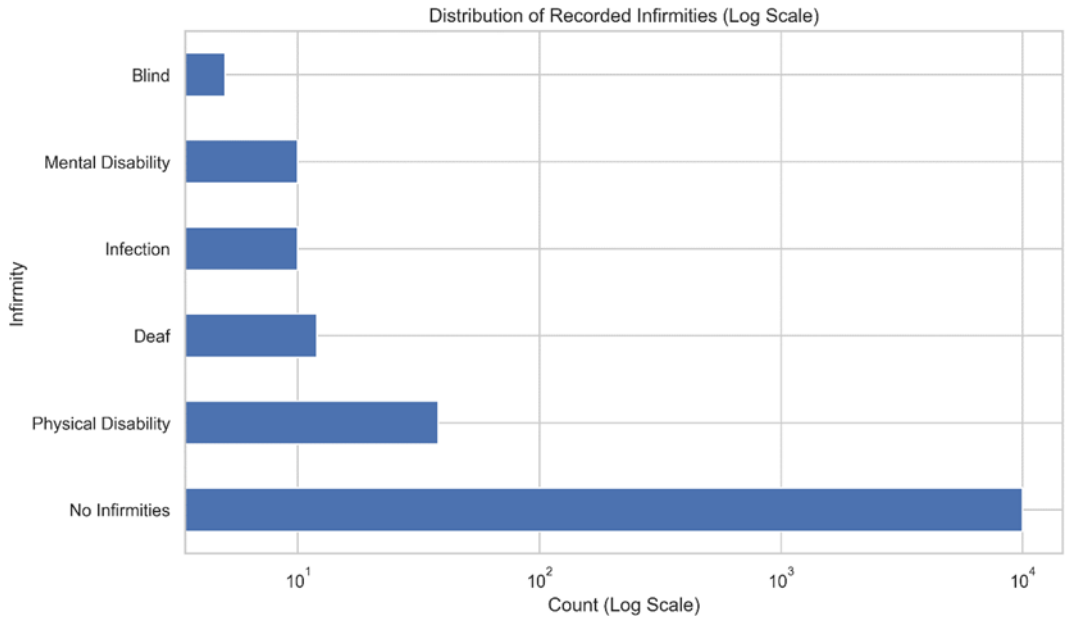


Figure 9: Religion distribution

In the boxplot in figure 9, no religion has lowest mean age of the respondents, the mean age of the church of England adherents is the highest.

INFIRMITY

Less than one (1) percent of the population suffer from one form of infirmity or the other, with the percentage of physically disabled persons being about 0.38 %. The town does not have a significant number of sick people.



OCCUPANCY LEVEL

The total number of houses in the town is three thousand, one hundred and five (3,105). While the average occupancy count is 3 persons.

Sixty-four (64) percent of the houses in the town are underused i.e. less than the average number of persons occupy the homes, this might be due to smaller households or lower demand for larger homes. while thirty-six (36) percent of the houses are overused. i.e. more than the average number persons occupy these homes. These might be due to shared homes especially students living together or higher demand for homes in those areas,

POPULATION GROWTH

“Population growth rate compares the average annual percentage change in populations, resulting from a surplus (or deficit) of births over deaths and the balance of migrants entering and leaving a country”. ((cia.gov)).

The **crude birth rate** is the number of live births each year per thousand of the population, from the calculation, the number of births for the year was calculated from using children aged 0. this was divided by the population. the crude birth rate was approximately 13 births per 1000.

The crude death rate is calculated as the number of deaths each year per thousand of the population in an area. In this study, the **death rate** was calculated by estimating the

differential of the number of people in one age bracket compared toa younger one. The assumption being that the number of people in each of these groups was the same when they were the same age. The annual death rate from the age band 60-64 when the population started reducing consistently is shown below and the total death rate was calculated as 9.085 per 1000.

	Age Comparison	Death Rate per Annum
0	60-64 vs 65-69	29.4
1	65-69 vs 70-74	9.8
2	70-74 vs 75-79	19.4
3	75-79 vs 80-84	7.6
4	80-84 vs 85-89	11.8
5	85-89 vs 90-94	9.6
6	90-94 vs 95-99	2.0
7	95-99 vs 100-104	-0.2
8	100-104 vs 105-109	1.6
9	105-109 vs 110-114	0.6

Immigration rate was calculated on the assumption that the Lodgers that included visitors in the relationship to the head of house column migrated temporarily to the town, while emigration rate were calculated on the assumption that the differential in the number of divorcees between men and women is because of the divorced men emigrating from the town.

The population growth rate was estimated to be around 0.36 percent.

COMMUTTERS

There are quite a lot of professionals that live in this town and commute to work out of town, this includes, ` university students, scientists, and many other professionals. The estimated number of commuters is about six hundred and thirty-seven. This amounts to about six (6) percent of the population commuting to work out of town.

CONCLUSION

In conclusion, results from the mock census data show a working-class population with low unemployment rate and mild population growth.

Amongst the younger population there is greater aversion to having no religious affiliation. Hence, a religious building shouldn’t be a priority.

The average occupancy rate is three (3) per house, however, a high proportion of these homes are under-used,

The town relatively has a healthy population, the current health facilities should be able to handle any minimal increase, since the increase in population is not that significant.

In five to ten years the retirement numbers will increase, there might be need to put infrastructure in place like in health and old people care homes, however it is not an immediate necessity.

Investment in general infrastructure is also not an immediate necessity since there is not much pressure on the population.

RECOMMENDATION

In the light of the above conclusion, I recommend as follows:

With the increasing number of commuters, I recommend building a train station, this not only takes away pressure from the existing road infrastructure, but also serves as catalyst for economic growth and employment opportunities for the unemployed. Most importantly, it makes the town more accessible.

I also, recommend increased spending for all levels early schooling up to high school. This is due to the steady increase in population of school age children between age 0 and 19.

LIMITATIONS

There were no data on birth, death, commuters' records etc, hence the study is littered with assumptions and extrapolations that render this work purely academic.

REFERENCES

Church of England - Wikipedia

https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/divorce/bulletins/divorcesinenglandandwales/2021

https://www.theguardian.com/world/2022/nov/29/why-is-the-christian-population-of-england-and-wales-declining

Population growth rate (cia.gov)

https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/divorce/bulletins/divorcesinenglandandwales/2021#:~:text=year%20on%20year.-,Divorce%20rates,(opposite%2Dsex%20only).