TITLE: Examining the Census Population Data and Make Recommendation on Land Development and Government Investment

INTRODUCTION.

In every ten years, population census is taken in the United Kingdom with the most recent population census conducted in 20201.The main purpose of the census is to provide government with accurate statistics of the population to enable better planning, develop policies and allocate resources.

In this report, I will analyze and make informed decisions regarding the development of an unoccupied land and investment priorities for a small town, based on a made-up census dataset that shows what the town's population might be like today. The dataset was created using Python's Faker package and is designed to emulate the structure of the 1881 UK Census, with modern-day characteristics. The dataset used in this analysis is named "T1_A24census4.csv".

The main objective of this report is to assess how the town can effectively utilize a vacant plot of land for development, considering a range of potential options, including high-density or low-density housing, a train station, religious buildings, or an emergency medical center. Additionally, the report will examine where future investment should be directed to meet the town's evolving needs, focusing on options such as employment and training, old age care, schooling, or general infrastructure improvements.

To make these decisions, the report will involve a thorough data cleaning process to address any inconsistencies, missing values, and potentially erroneous data. This report will leverage the findings from these analyses to provide a data-driven justification for the most suitable development on the unoccupied land and recommend the area of investment that will have the most significant positive impact on the town's future.

Data Cleaning Process and Methodology.

The data consist of various columns such as: House Number, Street, First Name, Surname, Age, Relationship to Head of House, Marital Status, Gender, Occupation, Infirmity and Religion. There were missing values at the Surname, Relationship to Head, Marital Status, Infirmity and Religion columns which needed some cleaning. There were no duplicate values found in the dataset.

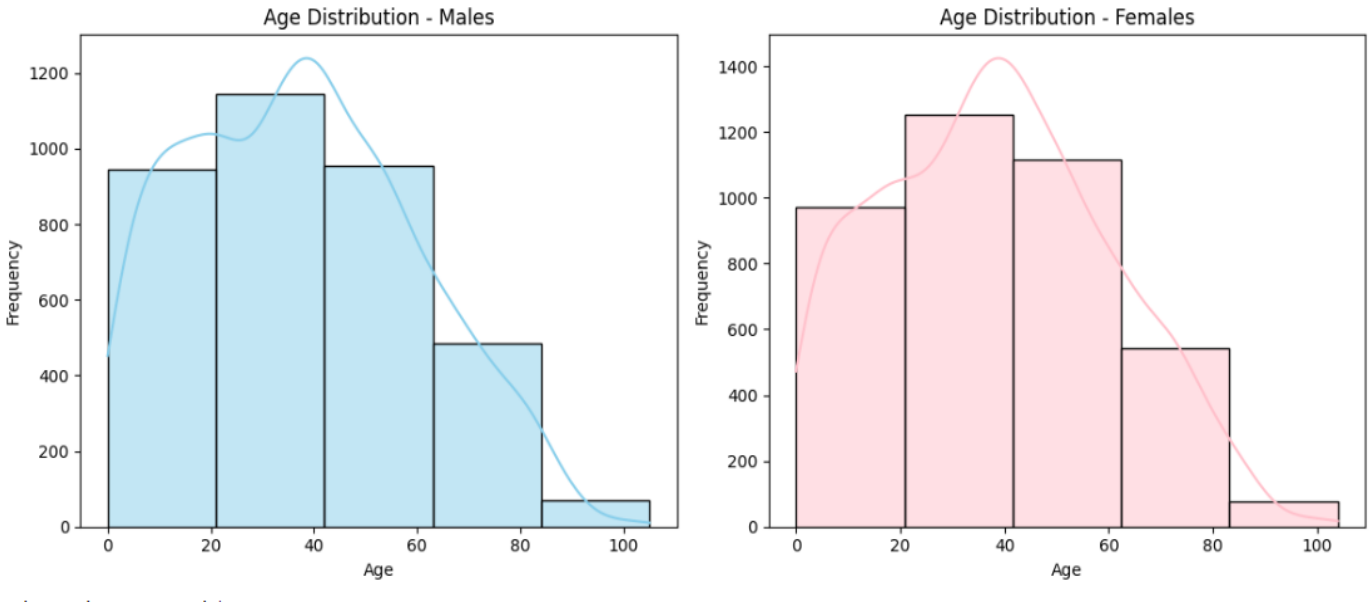
Methodology in Cleaning the "Religion" column: The missing values in the Religion column were replaced based on the condition that if the "Religion" of all the people at the same "House Number" & "Street" is same fill missing "Religion" value to that religion, if it is different mark the missing value "Unknown". Moreover, some "religions" values such as Dark brotherhood, Bahai, The Circle of Magi and Greek (obviously not a religion but a language) were replaced with "Unknown". The templars in the Religion column were replaced with Christians this is because, the templars can be traced to be Christians. They were Catholics and were closely tied to the Catholic church. Their beliefs and practices were aligned with Catholicism.

Methodology in Cleaning the "Marital Status" column: From the data, there are 1571 missing values in the marital status column. All these missing values were individuals who were below the age of 18 and regarded as minors. However, there are minors whose marital status were either "Married", "Divorced", "Widowed", "Single" in the dataset. These minors' marital status values were replaced with "N/A" in the Marital Status column.

Methodology in Cleaning the "Surname" column: The missing values in the "Surname" column were replaced based on the condition that if the "Surname" of all the existing individuals at the same "House Number" & "Street" are the same, fill missing "Surname" value with that "Surname", if different mark it as "Unknown".

The "infirmity" column was dropped. This is because the column will not be useful in the analysis.

Analysis on the Age Distribution of the Data.



The age distribution of both data appears unimodal, with one clear peak around the 30–40 age group. Both distributions are skewed to the right, meaning there are more younger individuals than older ones. The highest frequency is for ages in the 30–40 range, where the count approaches 1400 for Females and about 1300 for Males. The data spans from 0 to 100 years. Younger age groups (0–40 years) have higher frequencies, while frequencies decrease gradually for ages beyond 40 years. Both males and females show a decline in population above 60 years, but the decline is more gradual for females. This may reflect higher life expectancy for females, as the tail for females extends slightly longer than for males. This distribution represents a population where younger individuals dominate.

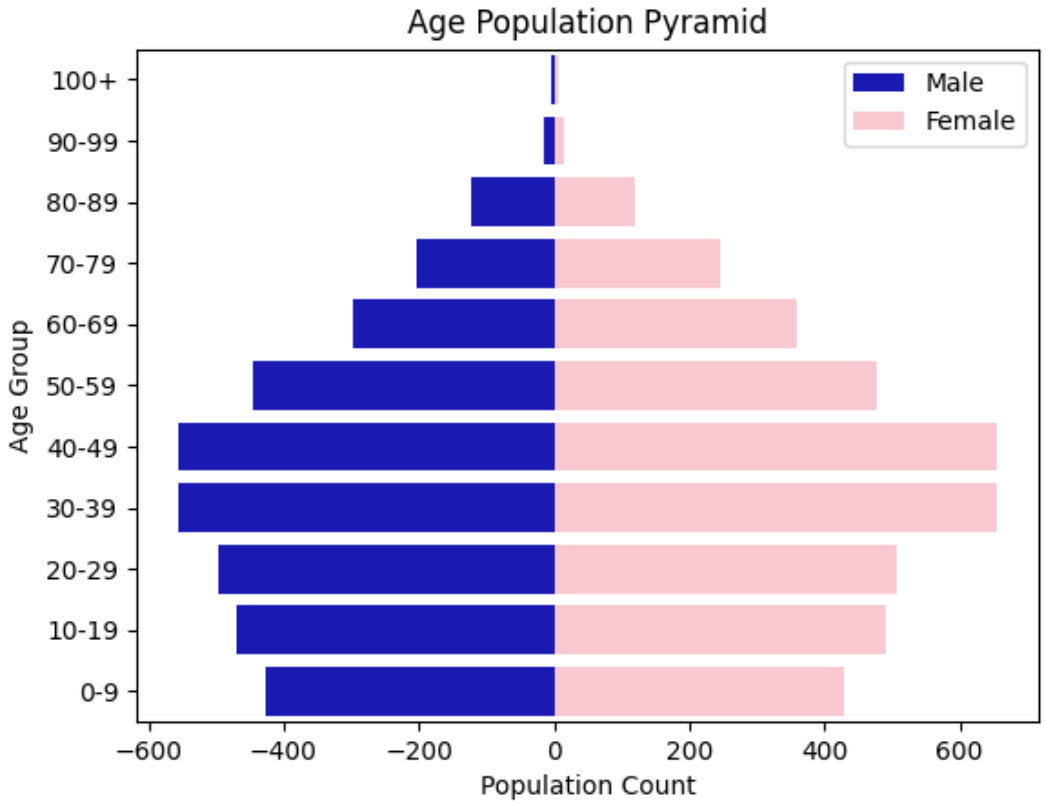


Figure 1

Figure 1 depicts the population pyramid of the data, indicating a balanced representation of males and females in most age groups. The shape of the pyramid wide at the base and narrows toward the top, consistent with a population that has more young individuals and fewer older ones. Males and females are relatively balanced in most age groups.

The pyramid exhibits a wide base (ages 0–19), indicating a relatively high birth rate. The middle portion (ages 20–49) indicates a huge number of strong working-age population essential for economic productivity. The top (ages 70+), representing older individuals in the population. The population significantly declines, typical of an aging population.

This demographic will require investments in education and healthcare. Aging Trends: The gradually increasing proportion of older adults (especially females) indicates the need to prepare for an aging population by focusing on elderly care and social support systems. Workforce: The strong 20–49 age group provides a demographic dividend, offering opportunities for economic growth if leveraged properly.

Analysis on the Religion of the Population.

After cleaning the "Religion" column, the actual (realistic) religions considered include Christian, Muslim, Hindu, Sikh, No Religion, and Unknown.

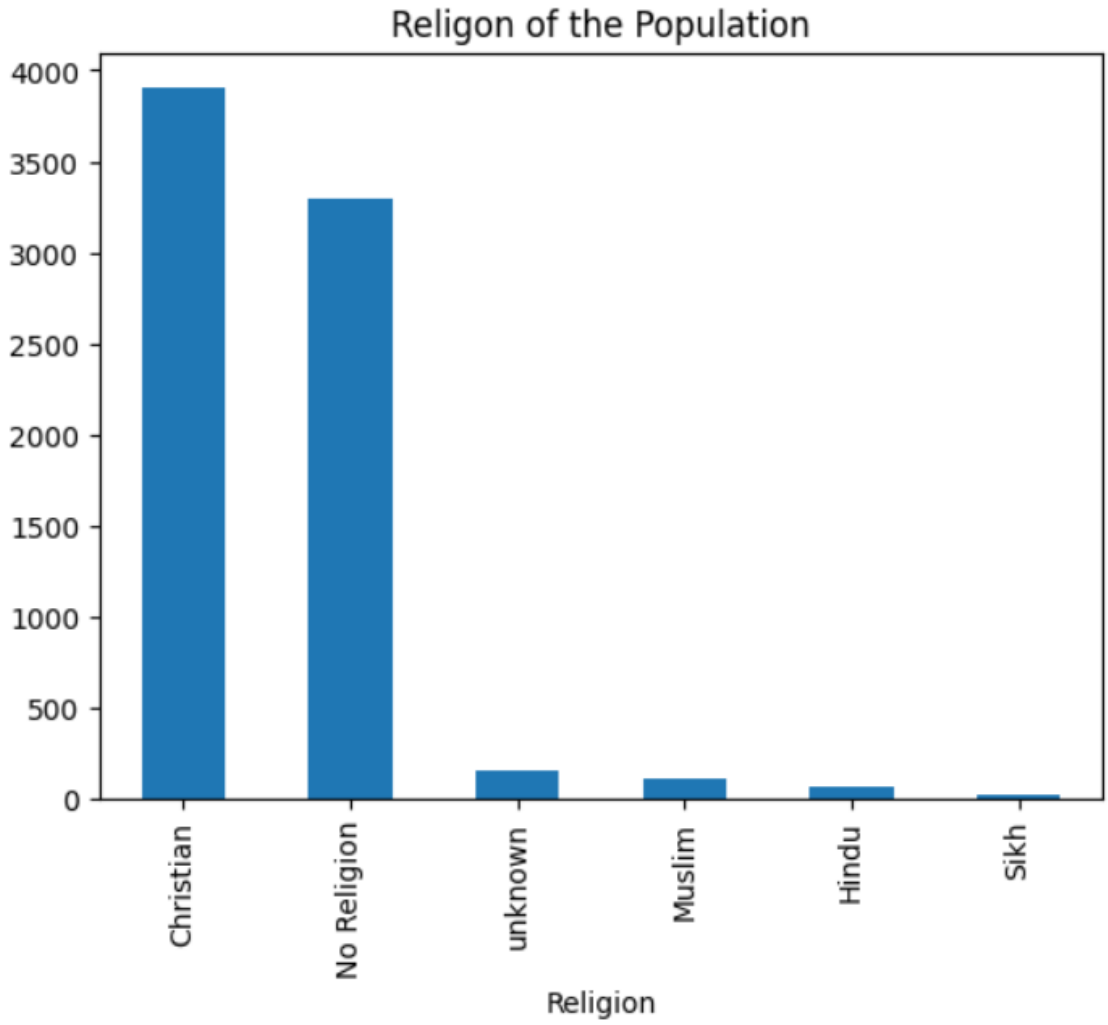


Figure 2

In *figure 2*, approximately 4,000 individuals of the population are Christians which is the highest religious group in the population. Individuals with no religion follows closely, who are slightly fewer individuals than Christians (around 3500–3900). Other religions like Muslim, Hindu, and Sikh have significantly lower representations, all below 500 individuals. There is an unknown category representing lack of data or unclassified individuals.

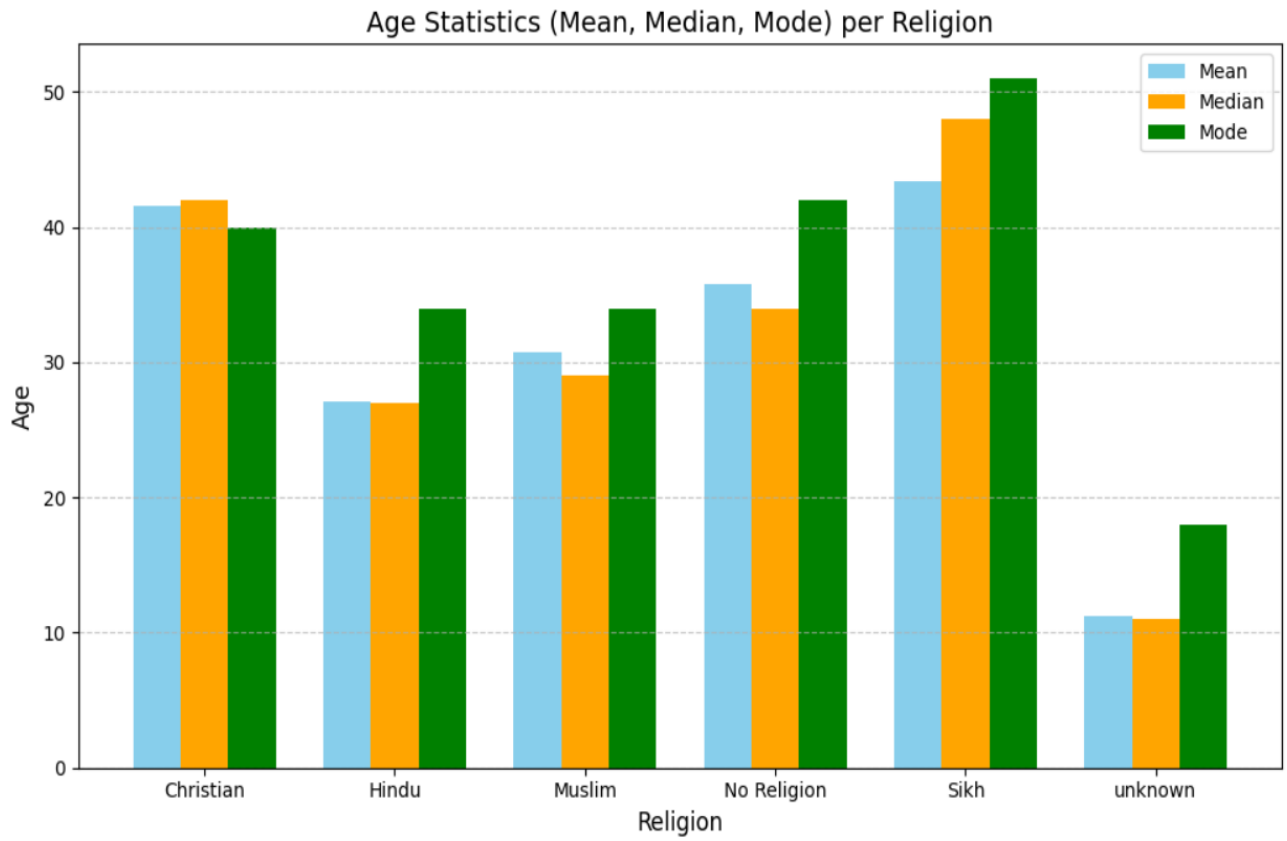


Figure 3

Figure 3 depicts the age statistics (Mean, Median, and Mode) for individuals categorized by their religion. Here's an analysis of the chart:

General Observations:

1. **Christian**: The Mean, Median, and Mode are closely aligned, indicating a symmetrical age distribution. The values are relatively high, around 40 years.
2. **Hindu**: The Mean and Median are lower compared to other groups (around mid-20s). The mode is higher than the Mean and Median, suggesting a skewed distribution with more individuals concentrated at an older age.
3. **Muslim**: All three metrics are similar, indicating a balanced age distribution. The values are in the low-30s range.

4. **No Religion**: The Mean, Median, and Mode are very close, suggesting a normal age distribution. The values are slightly higher than those of Muslims, around mid-to-high 30s.
5. **Sikh**: This group has the highest values for Mean, Median, and Mode, all exceeding 45. The similarity across metrics indicates a consistent and possibly older age demographic.
6. **Unknown**: The Mean and Median are low, while the Mode is significantly higher. This suggests that while the majority may be younger, there is a notable number of individuals at a specific older age less than 20 years.

Key Observations:

- The Sikh group stands out as having the oldest age demographic, followed by Christians.
- Hindu individuals seem to represent the youngest demographic, with a considerable gap between their Mean/Median and Mode.
- The "Unknown" category largely shows individuals below the age of 20 years. This could be that these individuals are not of religious dedicated age to really decided which religion they truly belong.

Analysis on the Occupation of the Population

In the Occupation column, I was able to identify the working class, unemployed, child, university student, students and pensioners or retirees in the town. Pensioners are individuals who are retired from active work. The working class are individuals who are actively working. Students consist of young individuals who are in school but not the university.

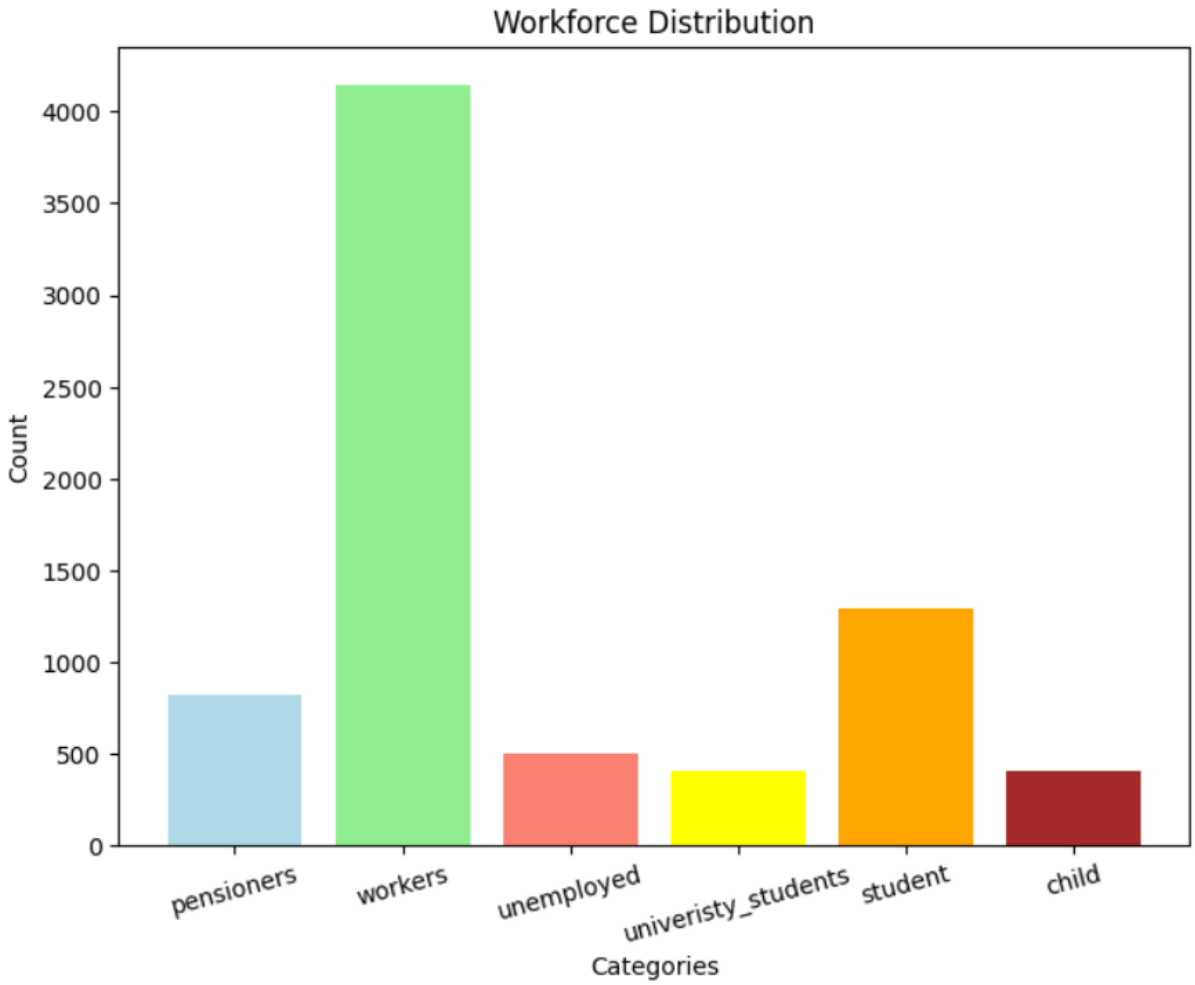


Figure 4

In fig 4, the working class or active workers in the population is about 4,146 which represents 54.9% of the population which is the highest group of people in the population. The number of retirees is about 820 representing 10.08% of the population. Student (1290) and University students (405) combined forms a significant part of the population (22.4%). Unemployment and

child counts are relatively low in the population of the town suggesting a balanced workforce and possibly a smaller portion of young dependents.

Identifying Commuters in the town: All university students (404) are commuters since there are no universities in the town. Also, individuals whose occupation are Pilot Airline (9), and Airline pilot (5) can be considered as commuters too since the town is a small-town sandwich between two large towns, it is very likely the do not have an airport and those two large towns have an airport hence the need to commute to those towns. In total, there would be 418 commuters in the town out of the total population of 7558 representing 5.53 % of the population.

The Death and Birth Rate of the Town

Considering the number of children aged zero (65) and the number of women aged 25-29 (assuming women in this age group are more likely to be fertile to give birth) gives birth to 268 children, The birth rate of the population is 24,253.73 per 100,000. Meaning the age group 25-29 is likely to produces 24,253.73 children out 100,000 births in a year representing 24.25% births. However, this assumption has some limitations: Fertility is not limited to the age group 25–29. The actual fertility window spans a broader range (e.g., 15–49 years). Some women aged 25–29 may already be past childbearing or not have children. Births might not be evenly distributed across all childbearing ages, making this estimate inaccurate. Fertility rates depend on social, economic, and cultural factors that vary widely.

The total death rate of the population is 20,497.24 per 100,000.00. A death rate of 20,497.27 per 100,000 means that for every 100,000 people in the population, about 20,497.24 people are expected to die which represents 20.5% over a certain period (usually a year).

The population growth rate of the town is 3,756.49. This is due the difference between birth rate and death rate, The difference between the birth rate and the death rate is positive, this means that the birth rate is higher than the death rate. This indicates a growing population, as more people are being born than are dying. Over time, this can lead to an increase in the overall population size.

RECOMMENDATIONS

I highly recommend that a Low-density housing should be built on the unoccupied plot of land. This is because, the age group of the population exhibits a huge number of strong working-age population. Majority of the population is actively working, which indicate a strong labor force. The working population is 4,146 representing 54.9% of the population which is the highest group of people in the population. These individuals earn income and may desire to start a family. Building a train station would not be efficient due to the low number of commuters in the town (418) out of which 404 are University students. Most of the population are Christians, and the town already have a place of worship hence no need to build a religious building. The number of individuals in other religious groups are very low and would not be efficient to build a place of worship for the other religious groups.

Furthermore, I recommend government invest in old age care centers in the population. The large number of the age group of the population exhibits a huge number of strong working-age population. There is evidence of strong labor force in the population. The working population is 4,146 representing 54.9% of the population which is the highest group of people in the population. This shows that soon, the number of retired people in the town will increase heavily and there would be a huge increase in demand for end-of-life care. Although employment and training are equally important, the unemployment in the population is significantly lower than the number of labor force in the population. Therefore, the need to prioritize an old age care to reduce the pressure on care facilities in the near future. The growing population of school-aged children is low hence increasing spending for schooling would not be efficient.

CONCLUSION

This report analyzed the town's census data to make informed recommendations regarding the development of a vacant plot of land and future government investments. The findings revealed a population dominated by working-age individuals and a significant number of younger residents, with a smaller but growing elderly demographic.

Based on these insights:

1. Land Development: Low-density housing is the most suitable option for the vacant land. This aligns with the high proportion of working-age residents who may seek stable, family-oriented housing options. Other options, such as a train station or additional religious buildings, are not recommended due to low commuter numbers and the already sufficient availability of worship space for Christians who are the majority.
2. Government Investment: The town should prioritize establishing old-age care centers. With a robust working-age population today, there will likely be a sharp increase in retirees in the near future, necessitating expanded end-of-life care services. Investments in schooling or employment programs, while beneficial, are less critical given the current low proportion of school-aged children and a strong labor force.

By addressing the town's immediate and future needs through targeted land development and investment, these recommendations aim to foster sustainable growth and ensure a balanced allocation of resources for the community.