Census Project Report

1. Abstract

This report undertakes an examination of a census conducted in a small town situated between two larger cities connected by a motorway. The primary objective is to conduct a comprehensive analysis of the population and explore the most effective utilization of an unused land parcel in the area. Initially presented as raw data fraught with errors, the census data poses challenges for analysis, necessitating the initial step of refining it into more usable and accurate data. This cleansing process will be addressed in the initial section of the report.

Subsequent sections will delve into the population analysis, with a focus on determining the optimal use of the vacant land. Commencing with demographics, the report will illustrate how the population is distributed in terms of age, followed by an assessment of whether population growth is on a positive or negative trajectory. Additional facets of the population to be scrutinized encompass employment trends, rates of divorce and marriage, commuting patterns to other cities, religious affiliations, and house occupancy.

2. Data Cleaning

To enhance the accuracy of the analysis, it was imperative to address errors present in the dataset initially. The initial observation revealed the presence of a redundant and irrelevant double index column, which was promptly eliminated from the dataset. Subsequently, an examination of attribute data types uncovered that all columns were categorized as object type, a designation unsuitable for attributes such as Age and Street number.

In the Age column, inconsistencies were identified in data entry, ranging from full integers to float entries and even a singular worded entry. These variations, compounded by the object data type, impeded mathematical and logical operations. To rectify this, the worded entry was transformed into a string integer, and all numerical entries were converted to float, followed by flooring to their lower integers. Post-flooring, the numbers were then converted to integers.

In standardizing the Marital Status column, efforts were made to unify entries, eliminating double entries such as using both "Single" and the letter "S" to represent the same status. Similar standardization measures were applied to divorced, married, and widowed entries.

Furthermore, instances of missing entries were identified in the Marital Status and Religion attributes. Upon closer examination of the religion column, it was determined that all missing entries pertained to minors under the age of 18. As a resolution, the marital status of these minors was imputed as 'NA.'

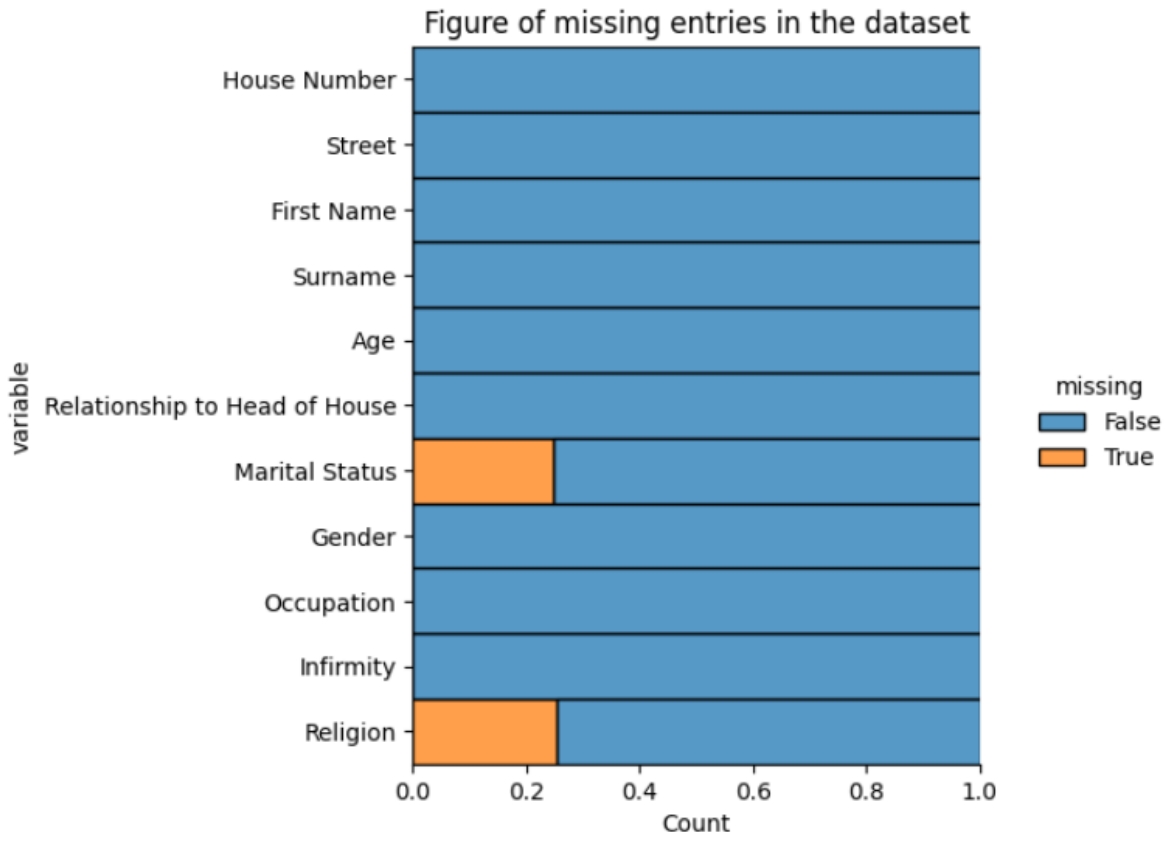

Figure 1: Missing figures per Attribute

In the Religion category, the majority of the missing entries were identified as belonging to minors. These entries were filled with 'Undecided,' reflecting the notion that individuals at this stage of life might not possess the rational capacity to make such decisions. Notably, it has been observed that many individuals undergo changes in their religious affiliations before the age of 24, sometimes even multiple times (Faith in Flux, 2020). The remaining 50 individuals without a specified religion were imputed with 'None' as their designation.

Erroneous entries in the Religion column were rectified, including instances where individuals stated 'Housekeeper' or 'Nope' as their religions, entries that lacked logical sense. Both 'Housekeeper' and 'Nope' entries were substituted with 'None.'

A household exhibiting numerous inconsistencies and illogical entries was excluded from the analysis. This household featured a 15-year-old girl as a widowed head of the household, and another where the head of the household was under 18 while the husband was 26. The head of the household status was corrected, transferring it from the wife (under 18) to the husband (of legal age).

To facilitate later analysis of retired individuals in the Occupation column, all entries indicating retirement in their professions were uniformly changed to 'Retired.' This simplification was deemed

practical for statistical operations, as retired individuals do not contribute to the active workforce of the city.

Other minor data cleaning measures were implemented, such as correcting the spelling of 'Niece' in the Relationship to Head of House column, rectifying a single entry in the house number from 'three' to '3,' and standardizing gender entries. Entries representing females as both 'Female' and 'f,' as well as males as 'Male' and 'm,' were unified to create consistent entries of 'Female' and 'Male,' respectively.

3. Demographic Analysis

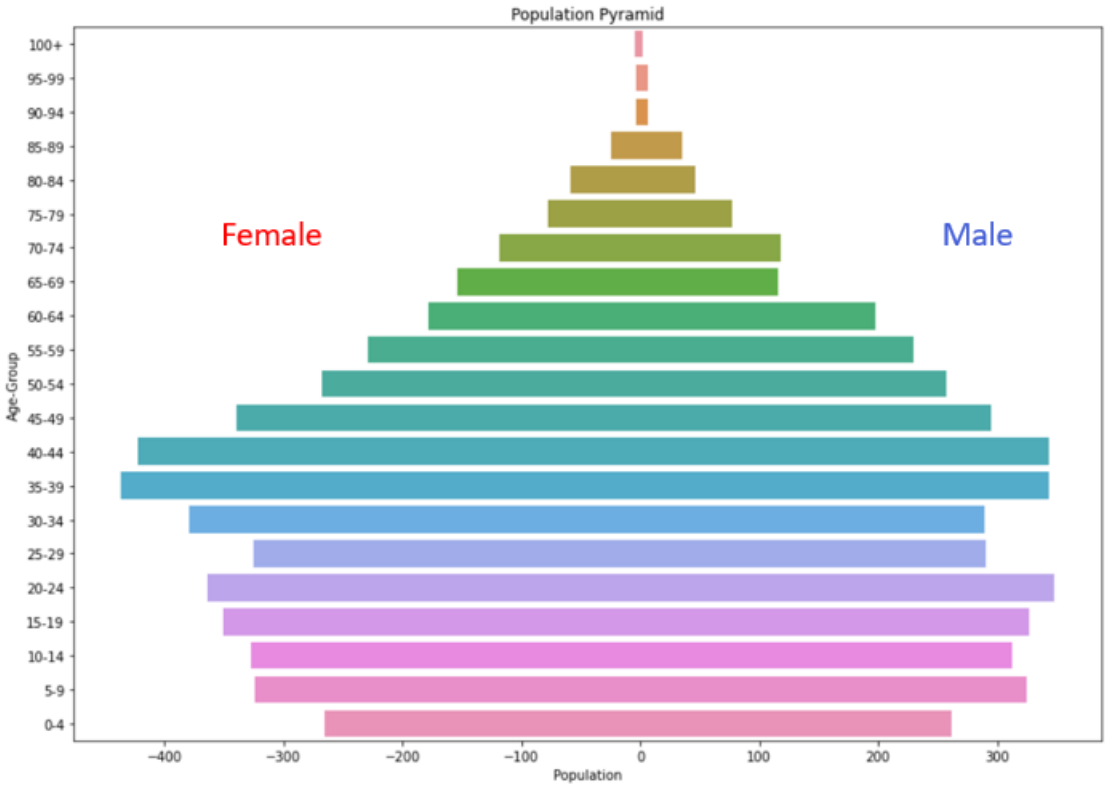

Figure 2: Population Pyramid

The demographic structure, as depicted by the population pyramid, reveals a notable predominance of adults (individuals aged 18 and above) in this small town, far outnumbering children or minors. The pyramid's narrow base signifies a low birth rate, suggesting a limited influx of new individuals into the population (Senguptra, 2014). This demographic pattern indicates a potential stagnation in the number of school-age children in the future.

The observed low birth rate in the small town may be attributed to the substantial participation of women in the workforce, surpassing the number of employed men in the region—2755 women compared to 2513 working men (Figure 3). Studies suggest that increased female workforce engagement correlates with lower childbirth rates due to the additional responsibilities women carry outside the household (Preshoff, 2014).

Furthermore, the town's high educational attainment is noteworthy, as research indicates a correlation between household education levels and the number of children. Highly educated couples tend to have fewer children, as evidenced by the calculated average of 2.19 children per household in the analysis (Jalovaraa, 2022). This underscores the impact of educational attainment on family size, with a higher level of education associated with a decreased likelihood of having a larger number of children.

		Count
Employment Status	Gender	
Employed	Female	2755
	Male	2513
Retired	Female	332
	Male	322
Student	Female	902
	Male	883
Unemployed	Female	350
	Male	211
University Student	Female	315
	Male	292

Figure 3: Employment Status Count per Gender

The expansive top section of the population pyramid indicates a substantial presence of elderly individuals in this region, suggesting an overall trend of increased life expectancy. Projections suggest a future surge in retirees, given the significant number of individuals currently within the workforce, as evidenced by the broad shape in the 30-60 age range of the pyramid. This working-class cohort is anticipated to contribute to the growing population of older individuals.

i). Affluence of Region

Having identified a substantial elderly population coexisting with a robust workforce, this town is indicative of affluence, boasting a balanced retirement age band and a workforce capable of financially supporting retirees through tax contributions. The workforce, characterized by a reduced number of children per household (as indicated by the pyramid's narrow base), results in fewer dependents, providing the working class with additional disposable income. This economic stability translates into substantial purchasing power, fostering investments and a high quality of life. To substantiate claims of affluence, the dependency ratio was calculated using the formula:

$$\frac{Number\ of\ Dependents\ or\ Non-Working\ Age\ Group}{Population\ aged\ between\ 15\ to\ 64(working\ age)} \times 100$$

was used to calculate the dependency ratio. The dependency ratio is a measure of the number of dependents aged zero to 14 and over the age of 65, compared with the total population aged 15 to 64 (Hayes, 2022). After calculation, it was found that the dependency ratio was 0.43 which is low. A low dependency ratio is one which is below 62.3 according to (Esri Data Development, 2022).

ii). Birth and Death rate

To validate the indications of a low birth rate in this region, further analysis was done by the help of calculation of the birth rate in the region. The crude birth rate is the number of live births occurring among the population of a given geographical area during a given year, per 1,000 mid-year total population of the given geographical area during the same year (UN. Statistical Office, 1991). The birth rate would be calculated according to the formula:

$$\frac{Number\ of\ live\ births}{Estimated\ year\ Population} \times 1000$$

From the calculations, it was found out the crude birth rate of this region was 10.59 births per thousand. This is a very low birth rate because low birth rates range from 10 to 20 births per thousand for a region (Arendt, 2016).

The death rate of the region cannot be calculated accurately since we do not have the number of people who died in this region for that year, but it can be estimated from the population pyramid and we find that from the broad peak of the pyramid, people are not dying at a rapid rate as they get older and the progression from the working age group to the old age group is not steep. This is showing that people are living longer and there is good public health for the region.

iii). Unemployment trends

A significant majority of individuals in this small town's labour force are employed, with 90 percent engaged in occupations. Consequently, the unemployment rate is relatively low, precisely 9.6 percent, determined by dividing the number of unemployed individuals by the sum of employed and unemployed people.

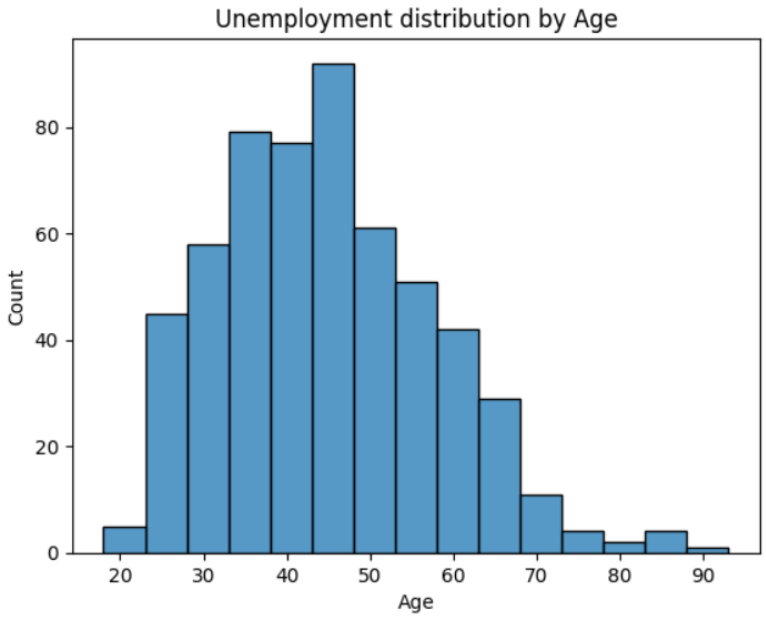

Figure 4: Unemployment distribution by Age

According to the Figure 4, most of the people who were unemployed were between the ages of 25 and 60. This is because that is the age band were people are economically active and they would have completed their education and will be seeking employment.

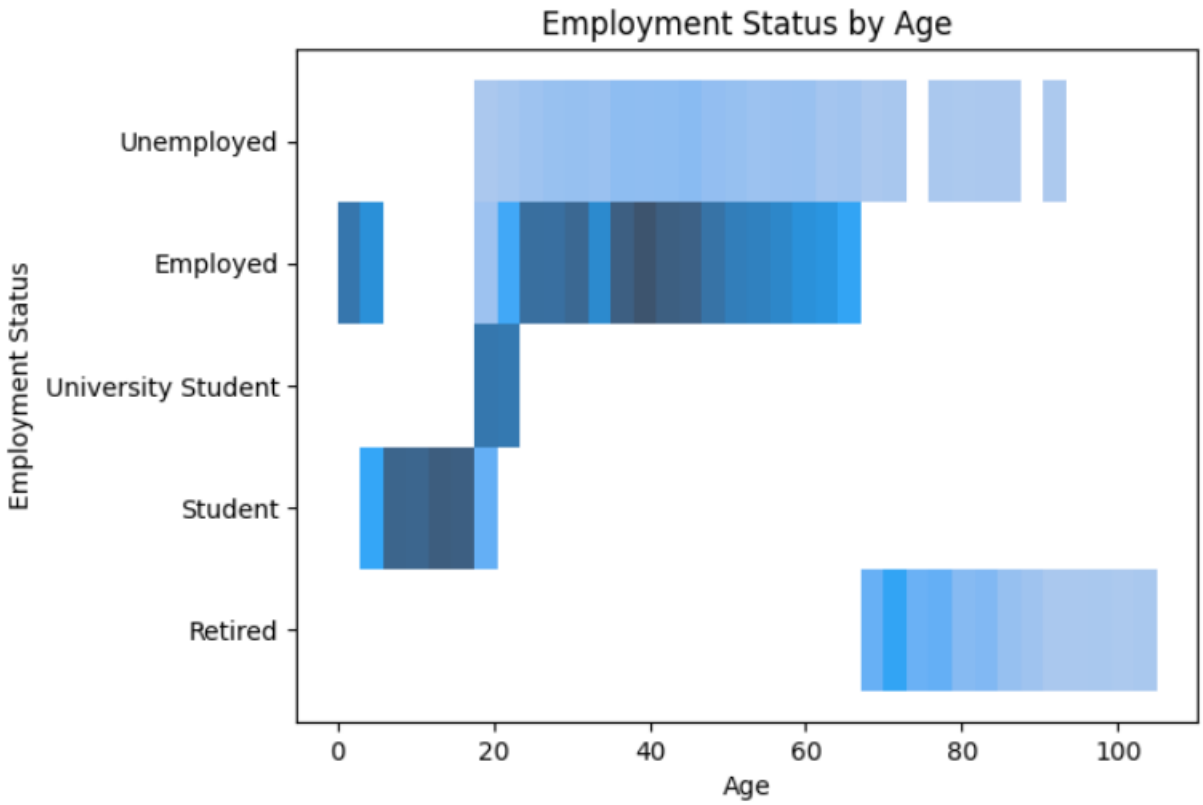

Figure 5: Occupations distribution by Age

Individuals under the age of 25 are likely to be actively engaged in education, as illustrated in Figure 5, where a significant proportion of university students fall within the 18 to 25 age range. The relatively low unemployment rates observed among those above 65 can be attributed to the majority being retired and no longer seeking employment. The remaining few may represent exceptions, as some individuals opt to continue working in their elderly years to stave off boredom or due to a genuine passion for their careers. The occurrence of unemployment among the elderly may stem from employers' concerns regarding their competence as they age.

iv). Religious affiliations

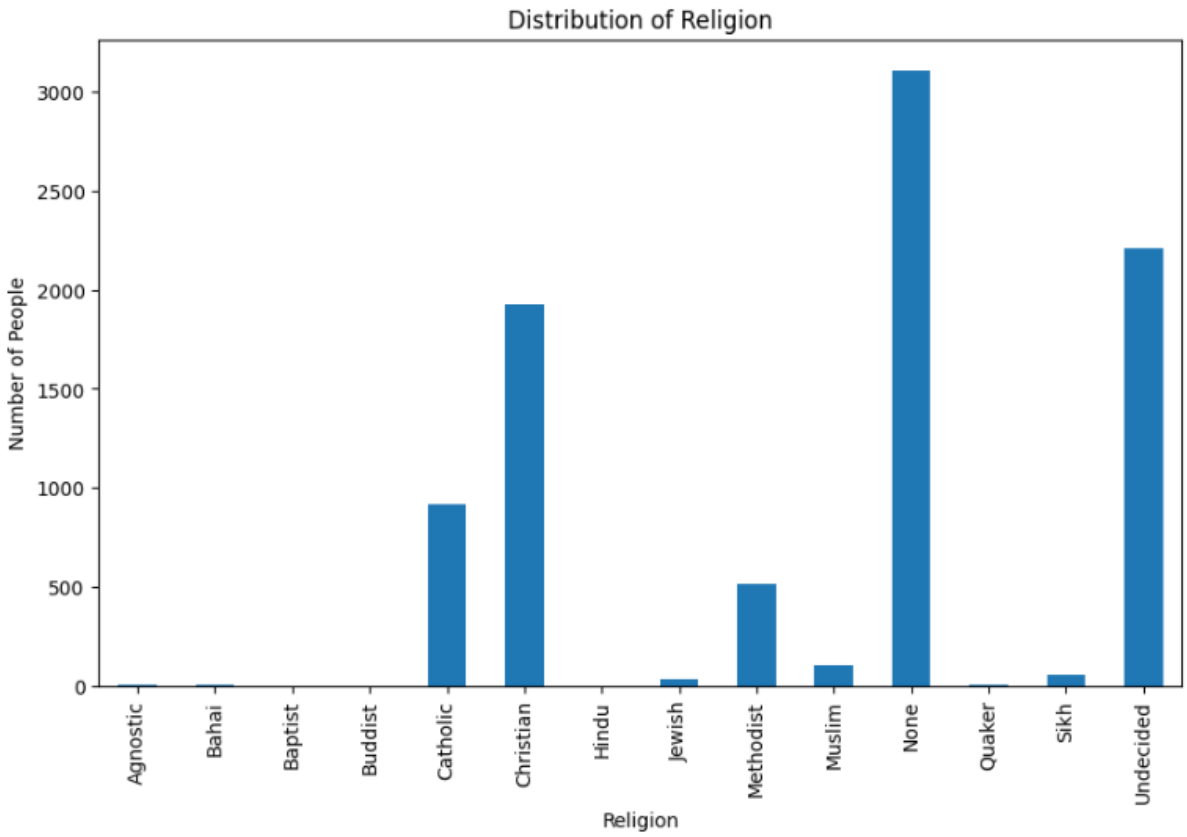

Figure 6: Distribution of Religion in the Region

From the distribution above, we can tell that most of the people who are religious in this area follow Christianity and Catholicism. A very good proportion of the people identified as not having any religion under the 'None' entry and a lot of minors under the age of 18 had missing entries in their religion hence being put as 'Undecided'. This made it difficult to predict religion transfer from parents to children as people have been seen to change their religions before the age of 24 too according to research.

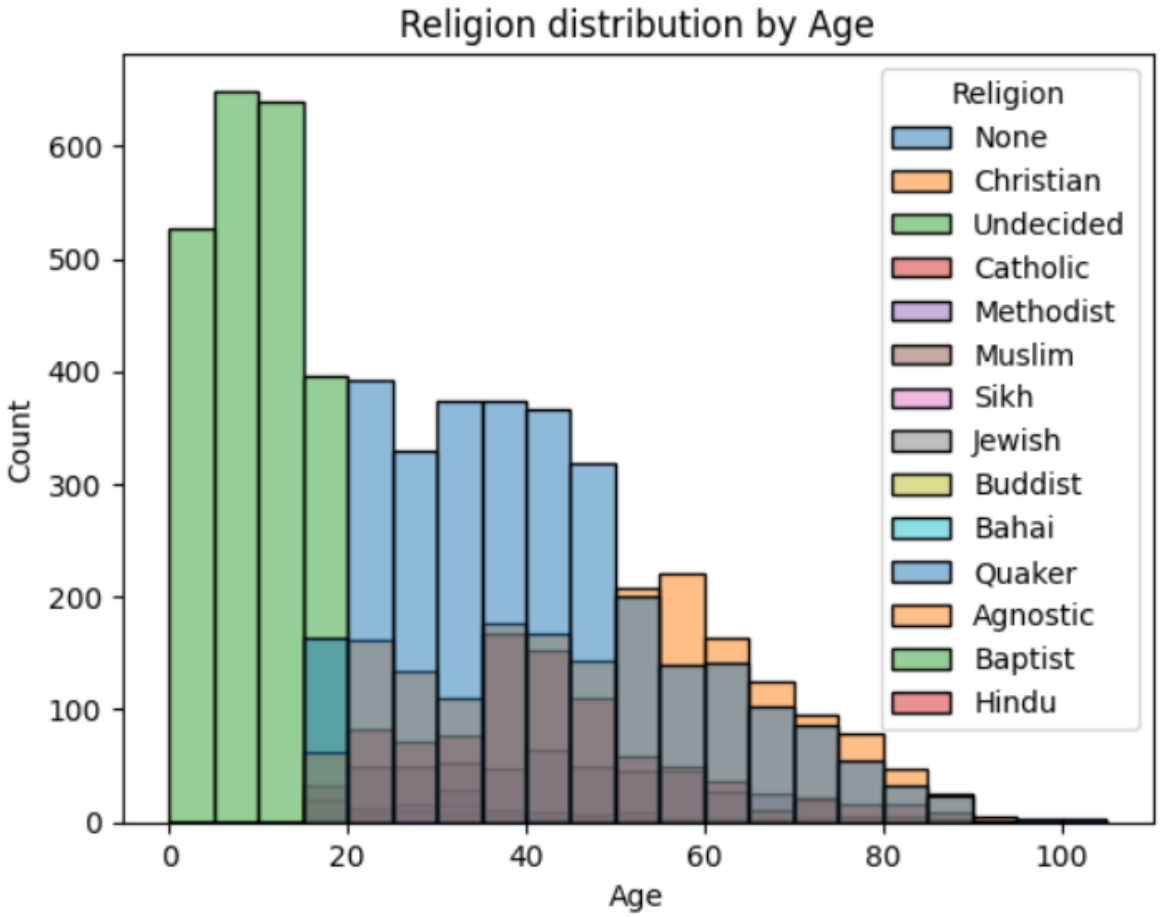

Figure 7: Religion distribution by Age

From Figure 7, we can tell that most people who identified as having no religion, are mostly people of the ages between 20 to 50. Most of the youth claimed that they did not practise or follow any religion. The more religious group were the elderly, people over the age of 50. There is only one religious building in this region, and it is for Catholics. Christians are the largest group in this region, but it would not make sense to build a church for Christians since most of the Christians are elderly and the following generations seem to be ditching the religion, the proportion of young people following Christianity is very low, hence it is a dying religion.

v). Divorce and Marriage

Value	Count	Frequency (%)
Single	3110	35.0%
Married	2424	27.3%
NA	2210	24.9%
Divorced	799	9.0%
Widowed	332	3.7%

Figure 8: Marital Status proportions

To find out how often people divorced in this town, a more refined approach had to be taken and it was to find the number of divorces per 1000 married women because it only included people who were exposed to the risk of divorce as opposed to crude rate which also included children. The divorce rate of the region was 652 people per thousand married women which indicates a very high divorce rate. The number of divorces in the region were seen to be also almost equally distributed between the males and the females. Marriage rate was calculated as 273 married people per 1000 population.

vi). House Occupancy Levels

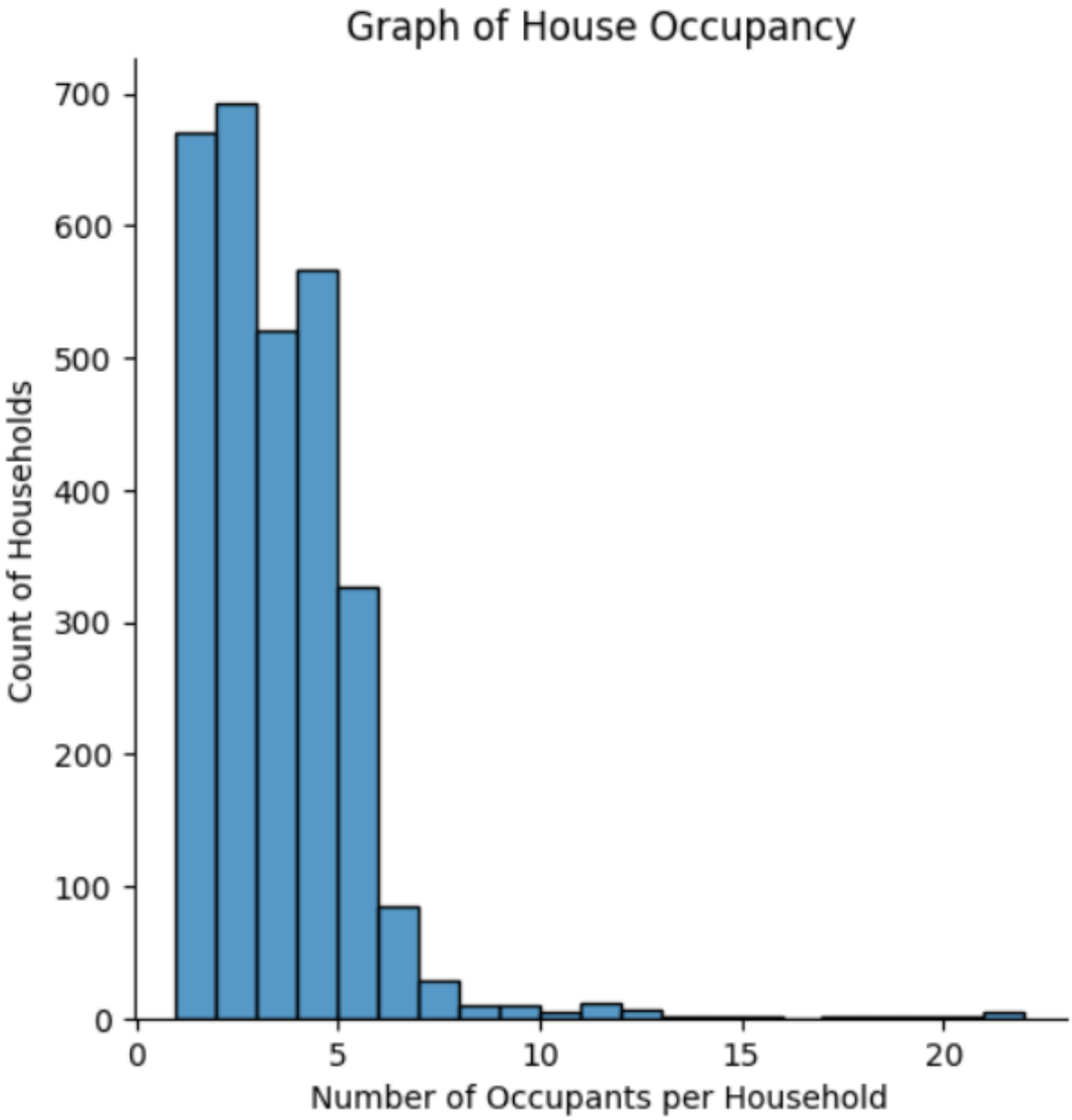

Figure 9: House Occupation distribution

The average number of people in a household in this town was found to be 3 occupants per household. With further investigation, it was found out that 46 percent of the houses had less than 3 occupants and

36 percent of households had over 3 people. With the average number of people per household in this region being 3 and more people living in households with 3 or less people as the majority, it shows that there is no overcrowding in houses and people are living comfortably. It has been found that in richer nations people tend to live in smaller numbers per household reducing the financial strain on the families and leaving them with more income (PopFacts, 2017). There is no underuse or overuse of houses in this region.

vii). Commuters

Most of the commuters in this region are University students who travel to nearby bigger cities by road for their education.

Value	Count	Frequency (%)
Student	1785	20.1%
Retired	654	7.4%
University Student	607	6.8%
Unemployed	561	6.3%
Child	527	5.9%
Administrator, sports	16	0.2%
Hotel manager	16	0.2%
Surveyor, mining	15	0.2%
Rural practice surveyor	15	0.2%
Management consultant	15	0.2%

Figure 10: University Students proportion

The proportion of these university students was found to be just 6.8 percent of the population which is quite insignificant for any considerations for construction of rail. Other professions which were more likely to be commuting to the other cities were Higher education lecturers and PhD students since there was no university in the region. Again, the 13 Higher education lectures and 11 PhD students could not improve the cost to benefit of constructing a railway since it will still be low.

4. Recommendations

Based on the conducted analysis, a prudent recommendation would be to utilize the unused land for the development of low-density housing. The region exhibits signs of a thriving economy, as evidenced by a substantial and employed workforce, coupled with a low dependency ratio. This demographic composition equips the working population with increased spending power, considering fewer financial obligations toward children and elderly individuals, thereby amplifying their purchasing capacity. Although the region is already affluent, significant economic expansion is anticipated, necessitating additional housing for the generally prosperous populace.

Furthermore, the data underscores a prolonged lifespan and high life expectancy among the residents, potentially attributed to a robust public health system and well-equipped medical facilities, contributing to an overall elevated quality of life. In anticipation of a growing elderly population, it is advisable to channel investments into old-age care, addressing the anticipated rise in the number of elderly individuals. As the economy continues to expand, more individuals are expected to experience financial prosperity, accompanied by an increase in life expectancy.

5. Bibliography

Arendt, H. (2022) *Birth rate*, *Wikipedia*. Wikimedia Foundation. Available at: https://en.wikipedia.org/wiki/Birth_rate#:~:text=Birth%20rates%20ranging%20from%2010,associated%20with%20low%20birth%20rates. (Accessed: December 7, 2022).

Author, F.in F. (2020) *Faith in Flux*, *Pew Research Center's Religion & Public Life Project*. Pew Research Center. Available at: https://www.pewresearch.org/religion/2009/04/27/faith-in-flux/ (Accessed: December 7, 2022).

Development, E.D. (2022) *How to use and interpret Esri's U.S. age dependency ratios*, *ArcGIS StoryMaps*. Esri. Available at: https://storymaps.arcgis.com/stories/eeb3d8b1dbef421fab9bff10019e5898 (Accessed: December 7, 2022).

Directorate, O.E.C.D.S. (2001) *OECD Glossary of statistical terms - crude birth rate definition*. Available at: https://stats.oecd.org/glossary/detail.asp?ID=490 (Accessed: December 7, 2022).

Hayes, A. (2022) *What is the dependency ratio, and how do you calculate it?*, *Investopedia*. Investopedia. Available at: https://www.investopedia.com/terms/d/dependencyratio.asp (Accessed: December 7, 2022).

Jalovaraa, M. (2022) *The highly educated often have two children - childlessness and high numbers of children more commonly seen among low- and medium-educated persons*, *Flux Consortium*. Available at: https://fluxconsortium.fi/the-highly-educated-often-have-two-children-childlessness-and-high-numbers-of-children-more-commonly-seen-among-low-and-medium-educated-persons/ (Accessed: December 7, 2022).

PopFacts (2017) *Population division* , *United Nations*. United Nations. Available at: https://www.un.org/development/desa/pd/ (Accessed: December 7, 2022).

Population pyramids: Powerful predictors of the future (2014) *YouTube*. YouTube. Available at: https://www.youtube.com/watch?v=RLmKfXwWQtE&t=287s (Accessed: December 7, 2022).

Senguptra, T. (2014) *Population pyramid*, *BuddingGeographers*. Available at: https://www.buddinggeographers.com/population-pyramid/ (Accessed: December 7, 2022).