**TABLE OF CONTENTS**

Contents

[1. Abstract 3](#_Toc121404154)

[2. Data Cleaning 3](#_Toc121404155)

[3. Demographic Analysis 5](#_Toc121404156)

[I. Affluency of Region 6](#_Toc121404157)

[II. Birth and Death rate 6](#_Toc121404158)

[III. Unemployment trends 7](#_Toc121404159)

[IV. Religious affiliations 9](#_Toc121404160)

[V. Divorce and Marriage 11](#_Toc121404161)

[VI. House Occupancy Levels 12](#_Toc121404162)

[VII. Commuters 13](#_Toc121404163)

[4. Recommendations 14](#_Toc121404164)

**TABLE OF FIGURES**

[Figure 1: Missing figures per Attribute 4](#_Toc121402840)

[Figure 2: Population Pyramid 5](#_Toc121402841)

[Figure 3: Employment Status Count per Gender 6](#_Toc121402842)

[Figure 4: Unemployment distribution by Age 7](#_Toc121402843)

[Figure 5: Occupations distribution by Age 8](#_Toc121402844)

[Figure 6: Distribution of Religion in the Region 9](#_Toc121402845)

[Figure 7: Religion distribution by Age 10](#_Toc121402846)

[Figure 8: Marital Status proportions 11](#_Toc121402847)

[Figure 9: House Occupation distribution 12](#_Toc121402848)

[Figure 10: University Students proportion 13](#_Toc121402849)

**Census Project Report**

# Abstract

This report conducts an analysis of a census on a small town located between two larger cities connected by motorway. The purpose is to make a thorough analysis on the population and see how best an unused plot of land in the area can be utilized. The census is firstly presented to us as raw data with a lot of errors which are problematic for analysis, so the first step would be to clean the raw data into more refined workable data. This will be covered in the first section of the report.

The sections that supersede the first one will focus on the analysis of the population with the aim of finding the optimum use of the vacant plot of land. This will start off with demographics to show how the population is distributed with respect to age and will go on to check if the population growth is following a positive or negative trajectory. The other components of the population that will be examined include employment trends, divorce and marriage rates, commuters to other cities, religious affiliations and house occupancy

# Data Cleaning

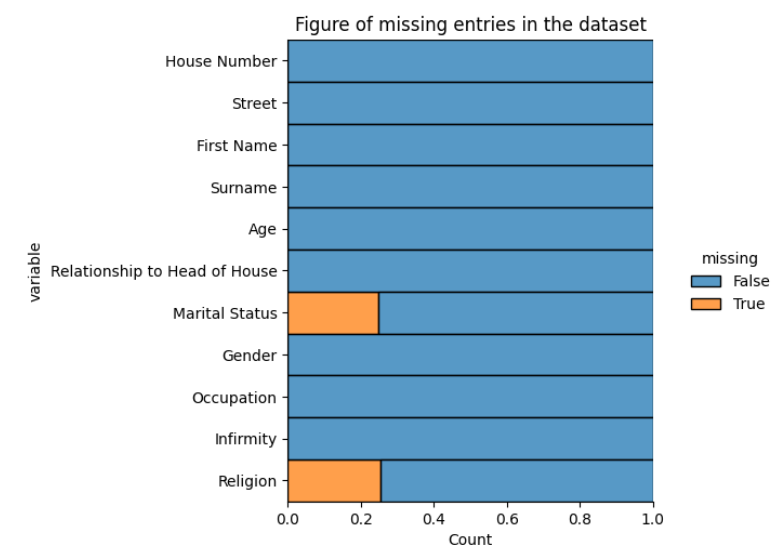
For the analysis to be more accurate, there had to be cleaning of the errors which existed within the data set first.

The first thing noticed was the double index which was serving as redundant and irrelevant data. This column was immediately dropped from the data set. After checking the data types of the attributes, it was found that all the columns were of object type and was not appropriate for others like Age and Street number.

There were inconsistencies in how the data was entered in the Age column, some entries had full integers as entries, some having float entries and one had a worded entry. The data was also in object form which would act as a hindrance to any mathematical or logical operations. To correct this, the worded entry was re-entered as a string integer and all the numbers were then converted to float and then floored to their lower numbers. After flooring the numbers, they were then converted to integers.

For marital status column, the entries were formalized into uniform ones instead of the double entries which were present for example, the “Single” entry also being represented by the letter “S” for other entries. The same was done for divorced, married and widowed.

There was evidence of some missing entries within the dataset amongst the Marital Status and Religion attributes. With further investigation of the religion column, all the missing entries were of the minors under the age of 18. The marital status of these minors was imputed to ‘NA’



**Figure 1: Missing figures per Attribute**

In Religion, most of the missing entries were also of minors. The entries were imputed by the entry ‘Undecided’ since they are people deemed not to be rational enough to be making this decision at this stage of their lives. Most people have been found to change their religions before the age of 24 and some even doing it more than once (*Faith in Flux,* 2020). The rest of the 50 people who did not have religion were imputed to ‘None’ as their entries.

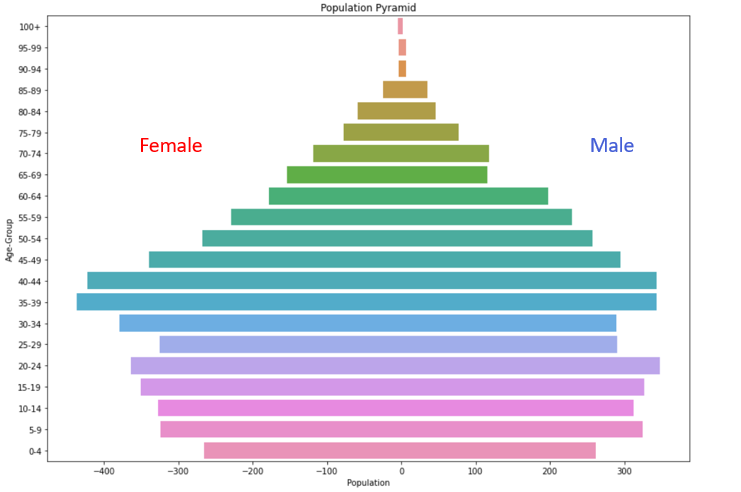
There were some wrong entries in the Religion column where people mentioned ‘Housekeeper’ as a religion they followed, which is an entry that doesn’t make sense. Other people had entered ‘Nope’ as entries. Both Housekeeper and Nope entries were replaced by ‘None’.

There was a household that was dropped which had too many inconsistencies and illogical entries. The household had a girl who was widowed at the age of 15 and was also the head of the house. There was another where the head of the house was under the age of 18 but the husband was 26 years old. The head of the house status was switched from the wife who is under 18 to the husband who is old enough.

To make it easier to analyse the retired people later in the occupation column, all the people who had retired in their professions in their occupation entries were changed to simply just ‘Retired’ without the rest of the occupation being mentioned because that information does no good for statistical operations since the people also do not contribute to the working force of the city.

Other minor data cleaning operations which were done include the correction of the spelling Niece in the Relationship to Head of House column, correcting a single entry in house number from ‘three’ to 3, correcting double entries of gender to single uniform entries in the Gender column. For example, Females were represented both as ‘Female’ and ‘f’. All were changed to just ‘Female’ and the same was done for Male entries.

# Demographic Analysis



**Figure 2: Population Pyramid**

The population pyramid shows that in this small town there are a lot more adults (people over the age of 18) than there are children or minors. From the narrow base of the population pyramid, it shows that there is a low birth rate in this town (Senguptra, 2014). The narrow base of the pyramid shows that there will not be an increase in school children in the future since not so many are being introduced into the population. The low birth rate in the small town could be a result of the involvement of women in the workforce as they even outnumber men who are employed in this region, having 2755 women as opposed to 2513 working men (Figure 3). The more women in a region are involved in the workplace, the less likely they are to have many children because of the responsibilities they have outside the household (Preshoff, 2014). The town must be highly educated too since there has been seen that there is a correlation between the number of children per household and the level of education in families (Jalovaraa, 2022). The more educated a couple is the, the greater the chances of them having less children indicated by the 2.19 children per household calculated in the analysis.

Graphical user interface, table

Description automatically generated

**Figure 3: Employment Status Count per Gender**

The broad shape at the top of the pyramid shows that there a lot of people in their old age. Generally, people are living longer in this region and projections show that there will be a lot more people in retirement in the future because of the significant amount of people in the current workforce indicated by the broad shape in the 30-60 age band of the pyramid. This working-class band of people will add on to the old age people in the future and increase the number.

## i). Affluence of Region

Having found out that there are a lot of elderly people who are living long in the nation with a good number of people in the working force, this shows that this is a rich town with a balanced retirement age band and a workforce to fund the retirees with the taxes that are paid. The working force also has less children per household indicated by the narrow base of the pyramid reducing the dependants that require money to be spent on them. This leaves the working class with more disposable income that gives them a lot of purchasing power to fund investments and live a good quality of life. To fortify the claims of this small town being affluent, the formula:

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:

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

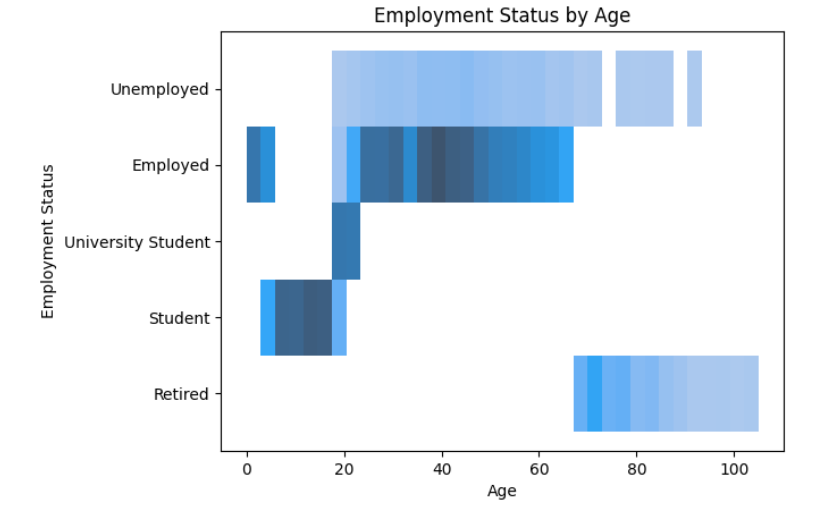
Most people in this small town who are seeking jobs and are in the labour force group are employed having 90 percent of them in occupations. This leaves 10 percent of the people unemployed, which is relatively low. The unemployment rate was found to be exactly 9.6 percent which was found by dividing the number of unemployed people by the sum of employed and unemployed people.

Chart, histogram

Description automatically generated

**Figure 4: Unemployment distribution by Age**

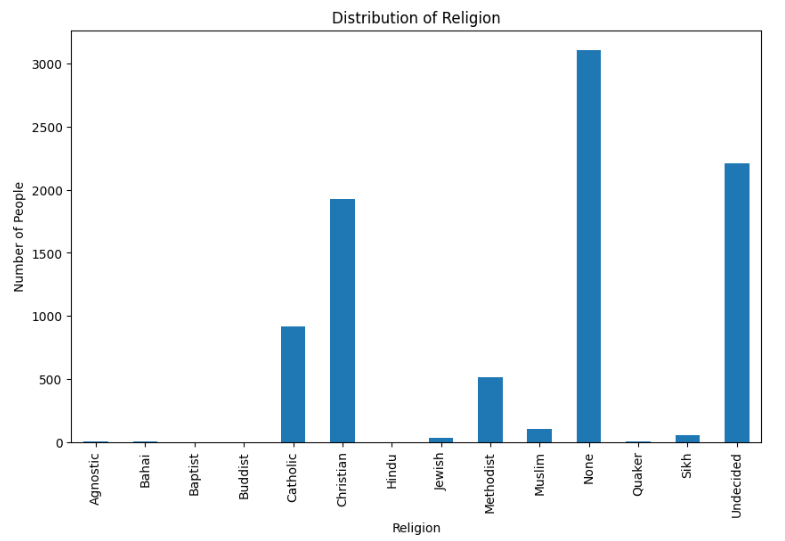
According to the Figure 4, most of the people who were unemployed were between the ages of 25 and 60. This is because that is the age band were people are economically active and they would have completed their education and will be seeking employment.



**Figure 5: Occupations distribution by Age**

People under the age of 25 will still be pursuing their education as we can see from figure 5 above that most of the University students are between the ages of 18 and 25. The low numbers of unemployment in people over the age of 65 is most likely because most of them are retired and are no more seeking employment. The few who are left should be an exception since people have been seen to still work in their elderly years in attempt to prevent boredom or simply for the love of their careers. The reality of elderly people who are unemployed could be a result of the questions employers have about elderly people’s competency as they get older.

## 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.

Chart, histogram

Description automatically generated

**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



**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

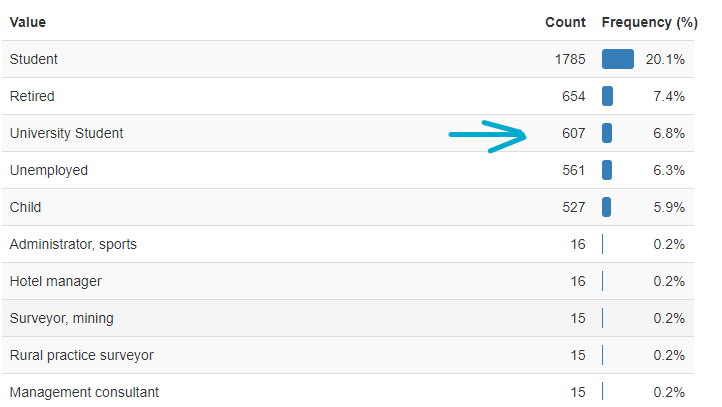
Chart, histogram

Description automatically generated

**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. 

**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.

# Recommendations

From the analysis made, it would be advisable for the unused plot of land to be used for the construction of low-density housing. There is evidence of this region being an expanding economy given by the large number of people in the work force who are also employed and the low dependency ratio which would give these people in the workforce more spending power since there are no too many children and elderly people to support with taxes to pay. The working class will be left with more money to spend and have a lot of purchasing power. It is already an affluent region but there is going to be considerable economic expansion and more housing will be needed for the generally affluent population.

It also has been evident that the people in this region tend to live long and have a high life expectancy possibly due to the good public health system and medical facilities contributed by a generally high quality of life. It would be advisable for investments to be poured into old age care to cater for the increasing old age population as the numbers are projected to increase with time. Since this is an expanding economy, more people are going to get richer, and the life expectancy of people will increase as a result.

# 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).