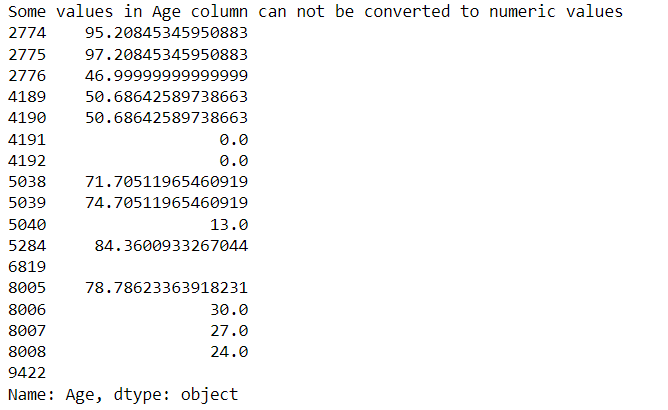
**Census Project Report**

This report analyses the census of a moderately sized town to make recommendations for investment in future services and use cases for developments on an unused plot of land. To make these recommendations, the census data has firstly been cleaned to correct data errors and missing records, which is detailed in the first section of this report.­­

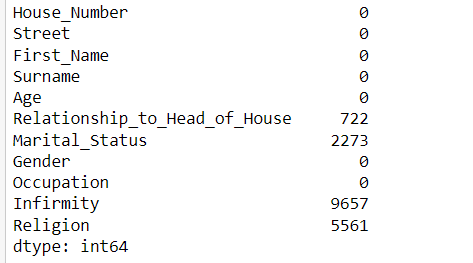
**Data Cleaning**

Age and House number must be integers but we got these data which are in the string type. So converting them to integer type is helpful for analysis.

****

In this data above some are empty strings and some of them are zeros and decimal values. Replacing these values based on the occupation and Gender of the person by taking the most frequent age of the people with same occupation and gender. There are no issues with converting the house numbers.

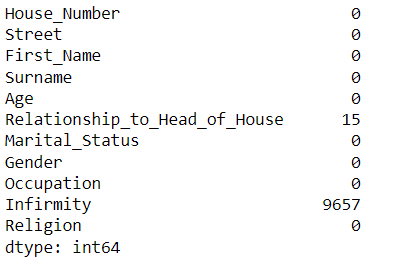
Now we try to know how many null values are present in each column.

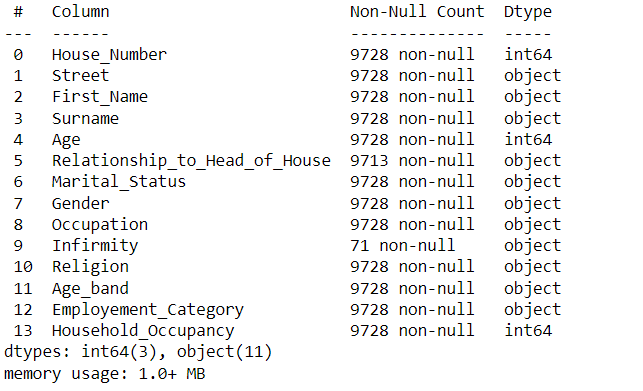


The null values in the column Relationship to head of house were reduced by making the criteria that If there is no the other person with same surname within the same house then we fill the column value as head of the house. Then the null values are reduced to 15.

For the persons with null values in the Marital status is given with the string None (The persons age less than or equal to 18 we are considering them as None(string)).

Religion of the people with the age less than 18 are filled with their parents religion.

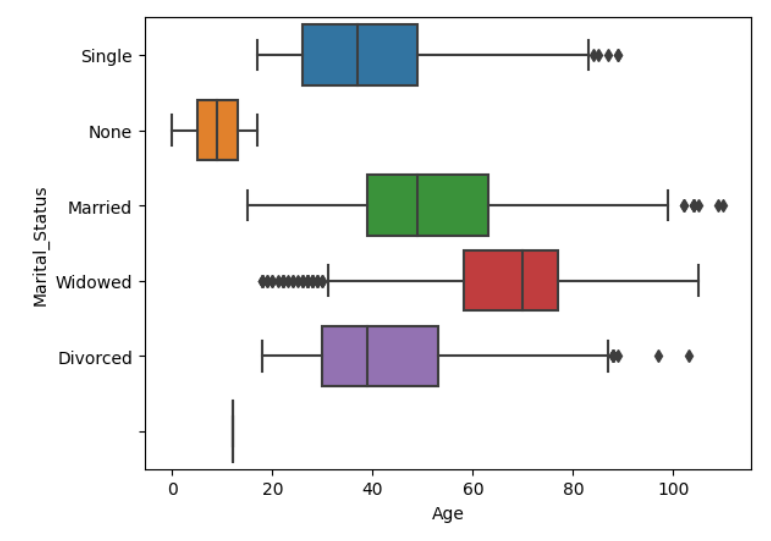


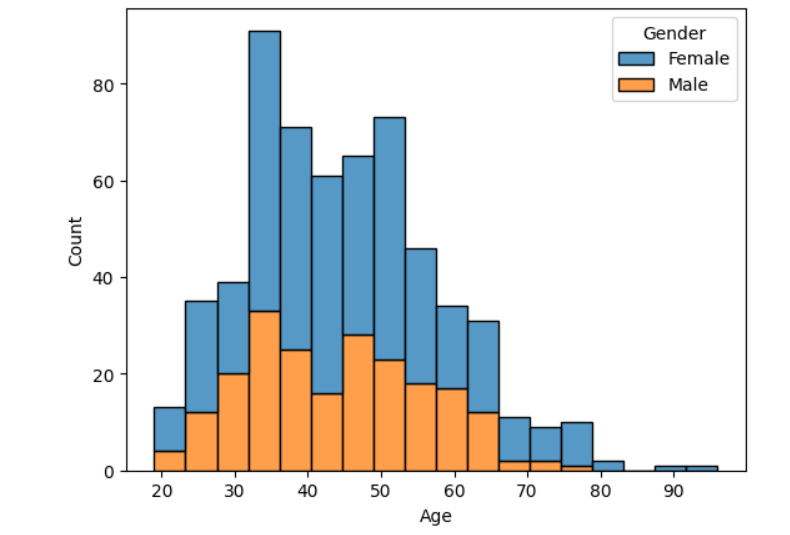


Here for the null values we have given the “None” string because of that the count is 9728.

NaN values in the marital status are given as “None” string if their age is less than 18, but for people with age > 18 the NaN values are left without any change because we don’t know whether they are married or not.

More percentage of unemployed people are female. People within the age band 30-60 have more number of unemployed people when compared to other ages.



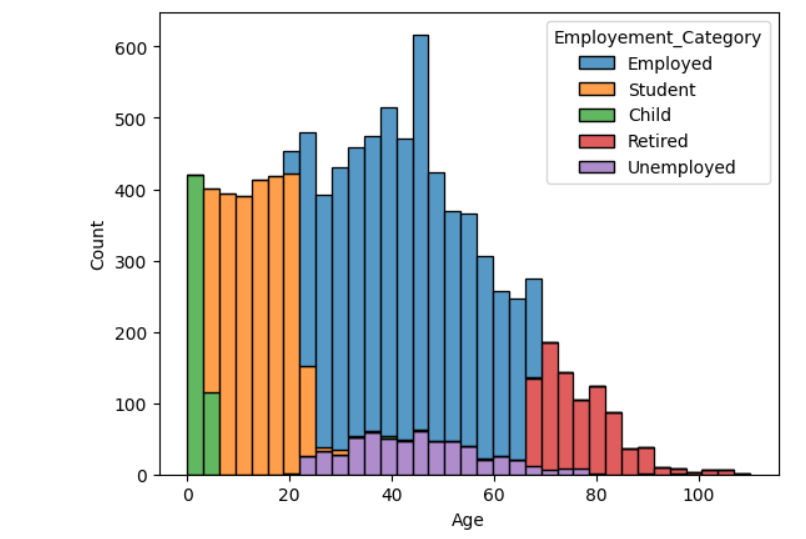


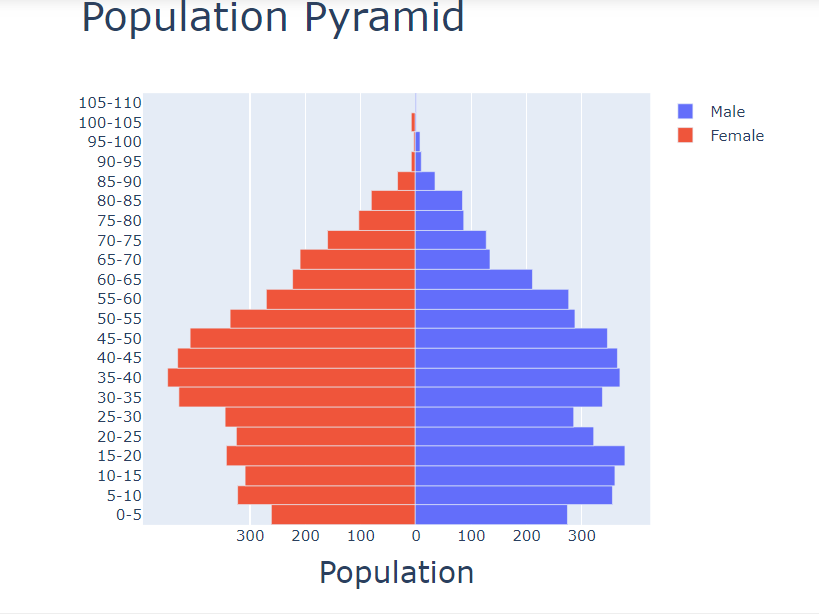
**Age band:** Ages placed into 5-year age bands for population pyramid.

**Employment category:** Simplified occupations with the values: Student (Child), Student, Employed, Unemployed, Retired.

**Household Occupancy:** a count of all individuals in a household. Note: this column should not be used for summation or counts.

The below plot show histogram of people with in age bands and different color shows different employment category of people. Most the population within the age band 20-65 are employed, 10-20 are students, above 65 are retired. Very less number of the total population are unemployed.





Current birth rate: Ratio of number of people with age 1 to total population multiplied by 1000

Current birth rate = 9.45

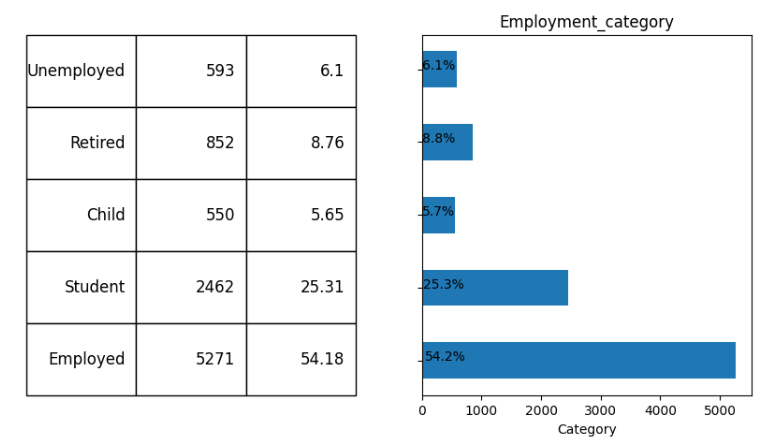
Birth rate 5 years back : Ratio of number of people with age 6 to total population multiplied by 1000

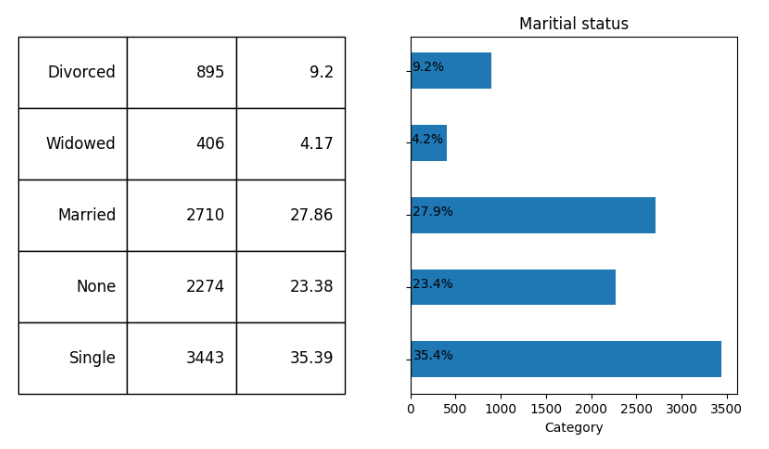
Birth rate 5 years back = 15.77

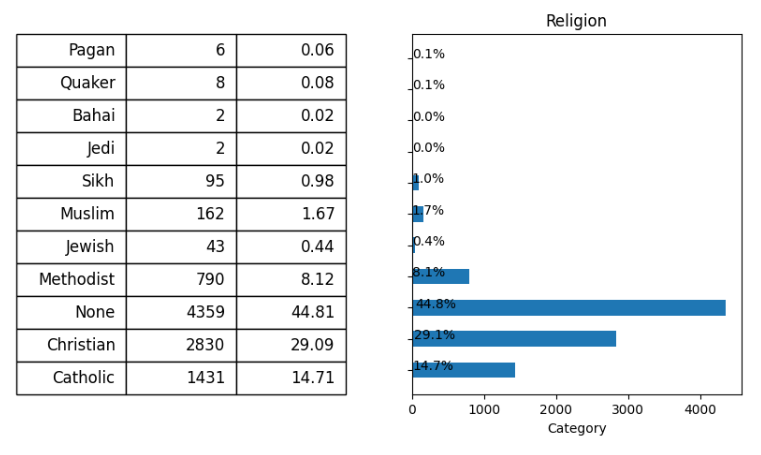
Death rate: calculated using Age band column. The death rate is calculated by estimating deaths by difference in age-bands for those over 65. Although there is decline across other age groups, the migration section details that these are likely individuals moving from the town (students or divorcees) as opposed to deaths. However, those over 65 are most likely to be retired and settled, and differences in these age bands are more likely to be deaths. Thus, by summing the difference of the left table for those above 65 and dividing by 5 (to account for the age-banding), the death rate is calculated as 8.5 deaths per thousand

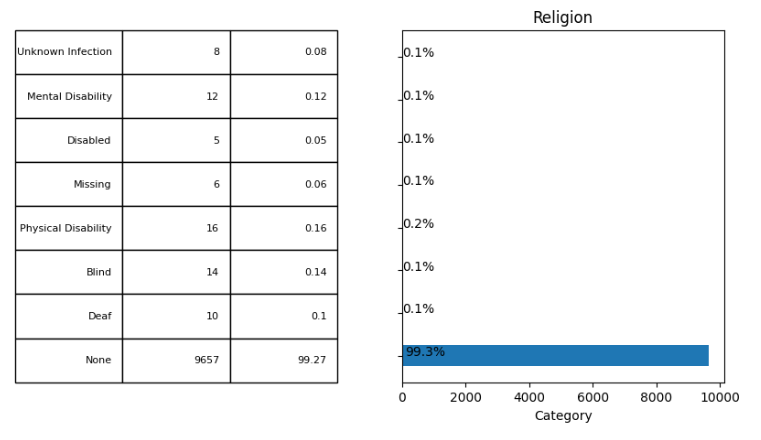
Death rate = 8.84

Population is shrinking and birth rate is decreased when compared to birthrate 5 years back.



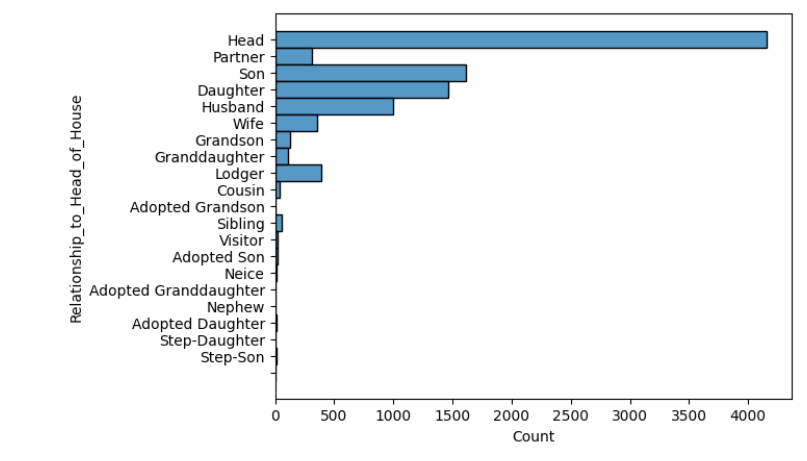




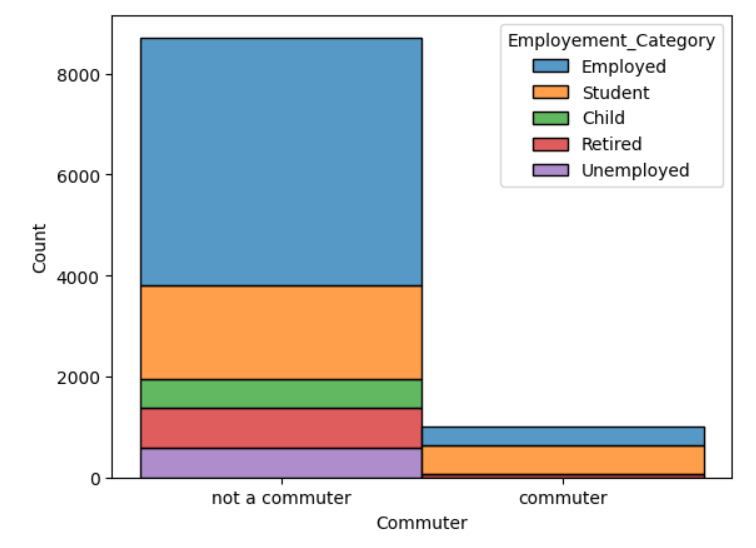


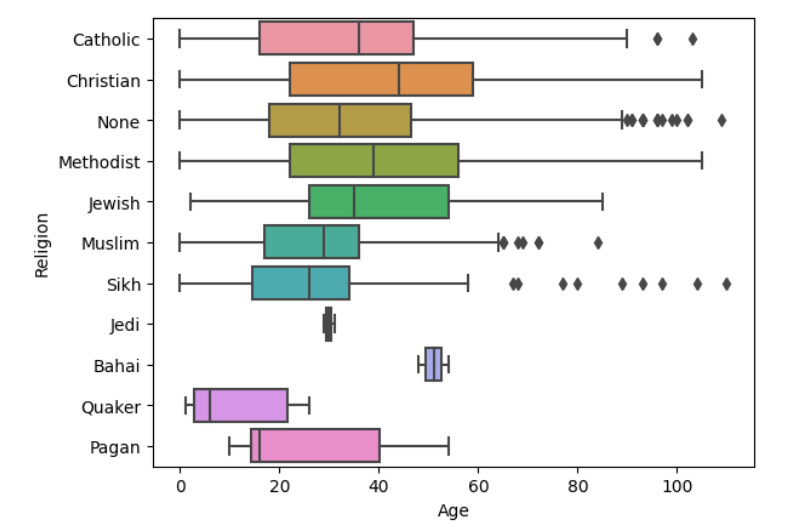
**(a)What should be built on an unoccupied plot of land that the local government wishes to develop?**

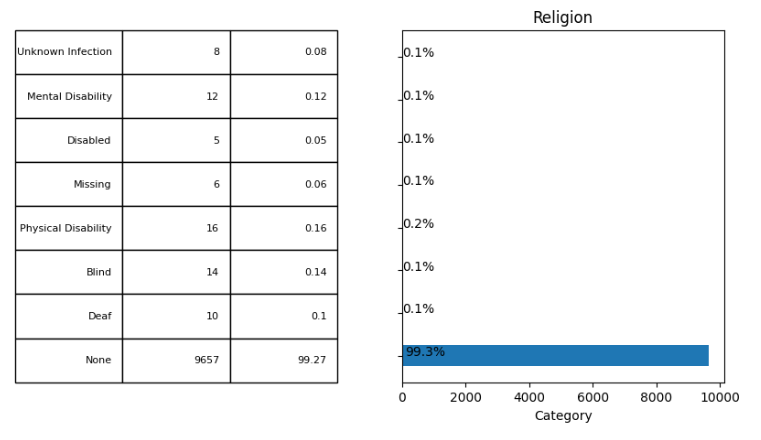
1. Percentage of people in lodger category is less and population is shrinking we need not build houses in unoccupied plot.



1. Percentages of commuters were also less if we consider only university students, PhD students and university professors and research scientists were the only commuters. So we need not build a train station.

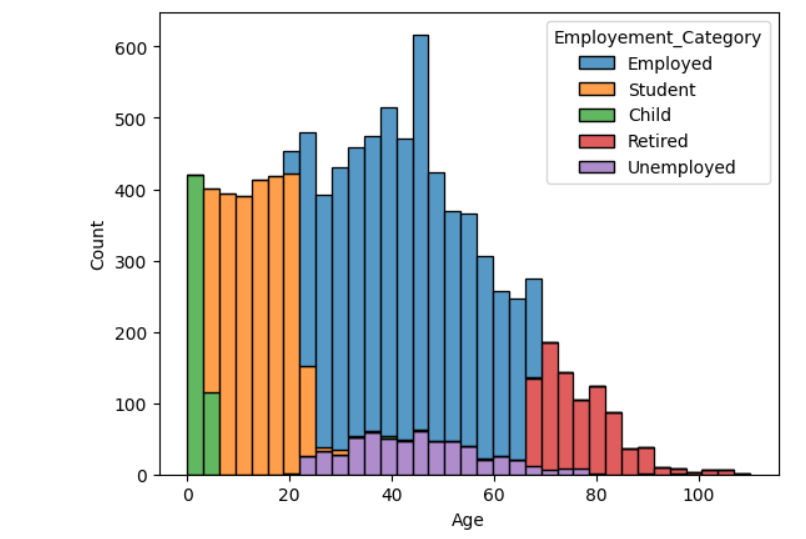


1. Christians were more than the Catholics so it is preferable to build a church in the unoccupied plot. 
2. Infirmity column values are mostly null so information about health conditions of people from the data is difficult to find so we can’t say anything about the Emergency medical building.



**(b) Which one of the following options should be invested in?**

1. Based on the data no of people falling under retirement category will increase in the following years when compared to increase in school going category.



Given the current Birth rate is 9.5 which is decreasing when compared with birth rate 5 years back and and death rate is 8.8

Increase in population of school going children will be close to birth\_rate \* population / 1000

In next 5 years no of school going children will be close to 5 \* birth\_rate \* population / 1000 + frequency of population within 0-15

expected\_change\_in\_students\_in\_5\_years = ((5 \*9.5\* 9728) / 1000) - 1 \* (dataset.Age\_band.value\_counts()["15-20"]) + (dataset.Age\_band.value\_counts()["0-5"])

**expected\_change\_in\_students\_in\_5\_years = 278.08**

current\_retired = len(dataset.loc[dataset.Employement\_Category == "Retired"])

retired\_in\_next5yrs = len(dataset.loc[(dataset.Employement\_Category == "Employed") & (dataset.Age >= 60)])

dead\_in\_next5yrs = (8.8 \* 9728 \* 5) / 1000

expected\_change\_in\_retired = retired\_in\_next5yrs + current\_retired - dead\_in\_next5yrs

**expected\_change\_in\_retired = 1019.96**

From the graph we can say percentage of people in unemployed category is less and from calculations above we can say expected increase of population in retired category is 1020 and expected increase in school going children is 278, increase in students can be even more less as the birth rate is decreasing , from the histogram we can see in future the more number of people will fall under retired category as currently employed category people are high in number.

**Final recommendations:**

Building church in the unoccupied land.

Money should be invested in old age homes.