# CENSUS PROJECT REPORT OF A MODERATELY SIZED TOWN


## Abstract

This is a report of census data of a moderately sized town sandwiched between two cities. The census data was being analysed to find out what should be built on an unoccupied plot of land which the local government wishes to develop and what the government should invest in. The data provided was explored, cleaned and formatted appropriately. Thereafter charts were plotted for visualization of the clean data to get the true position of the town in certain areas using certain criteria. Based on observations from the charts plotted, recommendations were made to guide the government appropriately.



WALLACE TUDEME

wallacepatrick2012@gmail.com

# Table of Contents

# List of Figures

# Introduction

## What is census and why is it important?

According to Wikipedia, "Census is the procedure of systematically calculating, acquiring and recording information about the members of a given population" (Wikipedia, Census - Wikipedia, 2021) and it is important because it provides required information for the central and local governments, health authorities, businesses and organisations to target their resources more efficiently with regards to providing education, healthcare, investment, social amenities and lots more (Office for National Statistics, n.d.).

## About our census data

This census report is based on a moderately sized town sandwiched between two cities. The data acquired from the census is going to be explored and considerations made based on insights from the explored data. The insights made from our data is to be used to advise the government on;

i. What should be built on an unoccupied plot of land that the local government wishes to develop?
ii. What should be invested in?

## Initial Data Exploration

Exploring the given data gotten from the conducted census, I discovered that our data had 11 columns namely;

- House Number
- Street
- First Name
- Surname
- Relationship to Head of House
- Marital Status
- Gender
- Occupation
- Infirmity
- Religion

Carrying-out initial data insights with a Pandas package called pandas profiling, some discoveries were made and are listed below;

i. There were 7,581 observations (This represents the number of persons that registered for the census).
ii. There were 3,398 missing cells which accounts for 4.1% of our entire dataset (This were the number of blank spaces).

Having discovered this, I decided to explore each column to see which rows were left blank. This gave rise to the need for data cleaning.

# Data Cleaning

Having explored the data, I had to explore each column individually using the unique method to get all the unique values in the data series. Thereafter I used the replace method to replace and adjust data as applicable.

## Cleaning the Age Data Series

I started with the Age Series and using the unique method, I found the values as shown in the picture below.

**Cleaning the Age Data Series**

```
In [6]: # To see the unique ages in the dataset

        original_census["Age"].unique()
```

```
Out[6]: array(['43', '44', '5', '25', '24', '3', '105', '73', '19', '10', '57',
               '30', '2', '88', '85', '52', '56', '28', '47', '20', '75', '59',
               '42', '61', '67', '66', '76', '77', '50', '26', '27', '55', '53',
               '29', '11', '22', '64', '68', '38', '34', '32', '37', '13', '69',
               '81', '60', '72', '31', '54', '83', '41', '39', '33', '4', '48',
               '62', '70', '46', '16', '84', '65', '15', '21', '40', '14', '12',
               '23', '74', '7', '51', '92', '6', '45', '63', '18', '36', '55.55',
               '17', '9', '71', 'Four', '49', '0', '80', '89', '91', '35', '1.75',
               'One', '58', '8', '1', '82', '79', '98', '100', '78', '94', '104',
               '96', '86', '87', '58.5088688', '0.5', '103', '93', '119', '90',
               '106', '55.52492408', '56.52492408', '131', '107', '55.27579403',
               '57.27579403', '99', '97', '-1', '95', '55.63583519',
               '59.22686093'], dtype=object)
```

*Figure 1 Picture showing unique values of unclean age data series.*

Seeing that the data in the age series was nowhere near being organised or in the right format, I had to make some adjustments to the data. The adjustments that were made are outline below:

i. I had to replace string containing texts “One” and “Four” to numerals “1” and “4” respectively.

ii. Afterwards, I converted the entire dataset to float for uniformity and then to integer as required because age is known to be in integer formats and there is no one that has an age 3.0 or 5.0. The correct format for such should be 3 or 5 respectively.

iii. Thereafter, the data was now looking a bit better with some anomalies though, as shown below.

```
# Check for unique ages

original_census["Age"].unique()
```

```
array([ 43,  44,   5,  25,  24,   3, 105,  73,  19,  10,  57,  30,   2,
        88,  85,  52,  56,  28,  47,  20,  75,  59,  42,  61,  67,  66,
        76,  77,  50,  26,  27,  55,  53,  29,  11,  22,  64,  68,  38,
        34,  32,  37,  13,  69,  81,  60,  72,  31,  54,  83,  41,  39,
        33,   4,  48,  62,  70,  46,  16,  84,  65,  15,  21,  40,  14,
        12,  23,  74,   7,  51,  92,   6,  45,  63,  18,  36,  17,   9,
        71,  49,   0,  80,  89,  91,  35,   1,  58,   8,  82,  79,  98,
       100,  78,  94, 104,  96,  86,  87, 103,  93, 119,  90, 106, 131,
       107,  99,  97,  -1,  95])
```

*Figure 2 Picture showing unique values of an average clean age data series.*

iv. I then proceeded to replace the age “-1” with integer “1” as no one is known to have a negative age and this must have been a typo error. I did this using the absolute function.

v. Lastly, I discovered that there was a woman named Josephine Holmes who entered her age to be 131 years and the oldest person known to be alive was 122 years old and obviously she has entered her age wrongly. She was married and retired. Knowing that she had a husband, I had to search for the husband’s information using some techniques (Further information on this can be gotten from my worksheet). Having gotten his information, I discovered that he was 107 years old. Knowing this I had to calculate the mean ages of all husbands and the mean ages of all wives

and calculated the difference between the mean values of both parties and I got the difference to be 3 years. I had to subtract 3 years form her husband's age of 107 years and I got 104 years which I used to replace the 131 years earlier declared by Josephine Holmes and this was done using the replace method.

Having done all this for the age data series, I got the oldest person in the town to be 119 years while the youngest was 0 year which are for new born in the town. The sorted cleaned age data is as shown below:

```
print(sorted(cleaned_age_data["Age"].unique()))
```

```
[0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 3
2, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 6
2, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 9
2, 93, 94, 95, 96, 97, 98, 99, 100, 103, 104, 105, 106, 107, 119]
```

*Figure 3 Picture showing unique values of cleaned age data series.*

## Cleaning the Religion Data Series

The religion data series was the next series that was explored using the unique method. I discovered that there were 15 unique entries namely;

- None
- NaN (used for minors)
- Christian
- Catholic
- Muslim
- Methodist
- Jewish
- Jedi
- Sikh
- Bahai
- Agnostic
- Undecided
- Blank spaces
- Baptist
- Private

Seeing all the unique entries, I had to research about the acceptable religions in the United Kingdom (UK) and I got to understand which religions were approved in the UK from Wikipedia (Wikipedia, Religion in the United Kingdom - Wikipedia, 2021). However, I could not find some religions in the publication which were entered as part of the Census religions. Further research on these religions showed the following;

i. Agnosticism is the view that the existence of God, or the supernatural is not known or knowable with any certainty (Wikipedia, Agnosticism - Wikipedia, 2021). By this, I came to the conclusion that Agnosticism is only a school of thought and not a religion; hence replacing all persons who filled in "Agnostic" as "Undecided" judging by the fact that they are yet to come to terms of believing in the existence of a supernatural being which is the foundation of religion. According to oxford online dictionary, "Religion is said to be the belief in and worship of a superhuman controlling power, especially a personal God or gods" (Oxford Languages and Google, n.d.).

ii. Jedi according to a BBC online publication, had a ruling against it by the Charity Commission in the UK. The Charity Commission ruled that Jediism which is the worship of the mythology of Star Wars, is not a religion (Jedi is not a religion, Charity Commission rules - BBC News, n.d.). By this, I had to replace all entries with "jedi" with "None" judging by the fact that they have no religion as stated by the ruling.
iii. I replaced "Blank spaces" with "Undecided". This is because if they had no religion or did not believe in religion, they would have filled in None. By not selecting any religion, I judged them to either be undecided about their choice of religion or undecided about being religious or not being religious.
iv. I replaced "Private" with "Undecided" as they would prefer not to state their religious stand.
v. I replaced all "NaN" values as "Undecided" as they were minors (aged below 18) and not qualified by law to make a decision about their religious preference and this they might do in the near future when they become adults (Office for National Statistics, n.d.).

Having done all this modification to our religion data series, the number of unique entries became 10 and they are as shown below along with their population count.

**Population of the various religions in the religion data series**

```
# Review of various element count in the Religion data series

cleaned_age_data["Religion"].value_counts()
```

```
4]: None         2624
    Undecided    1728
    Christian    1713
    Catholic      794
    Methodist     506
    Muslim        127
    Sikh           49
    Jewish         35
    Bahai           3
    Baptist         2
    Name: Religion, dtype: int64
```

*Figure 4 Picture showing unique values of cleaned religion data series with their population count.*

## Cleaning the Gender Data Series

For the gender data series, using the unique method I found 8 unique entries listed below;

- Female
- Male
- F
- M
- Fe-male
- MALE
- FEMALE
- Blank spaces

First, I had to check the entries with the blank spaces to find out why they were blank and what the right gender should be. I discovered that there were two entries one a son and the other a wife. I had to replace them with Male and Female gender respectively also I decided to use the case type of "Male" and "Female" for my unique gender entries. I had to ensure that there was uniformity in the format of my

gender series hence I assigned all entries to the above case type as applicable. This left me with just two unique entries Male and Female.

## Comprehensive Data Insights

From the data explored, charts and plots derived are outlined and implications explained as seen below;

### Age

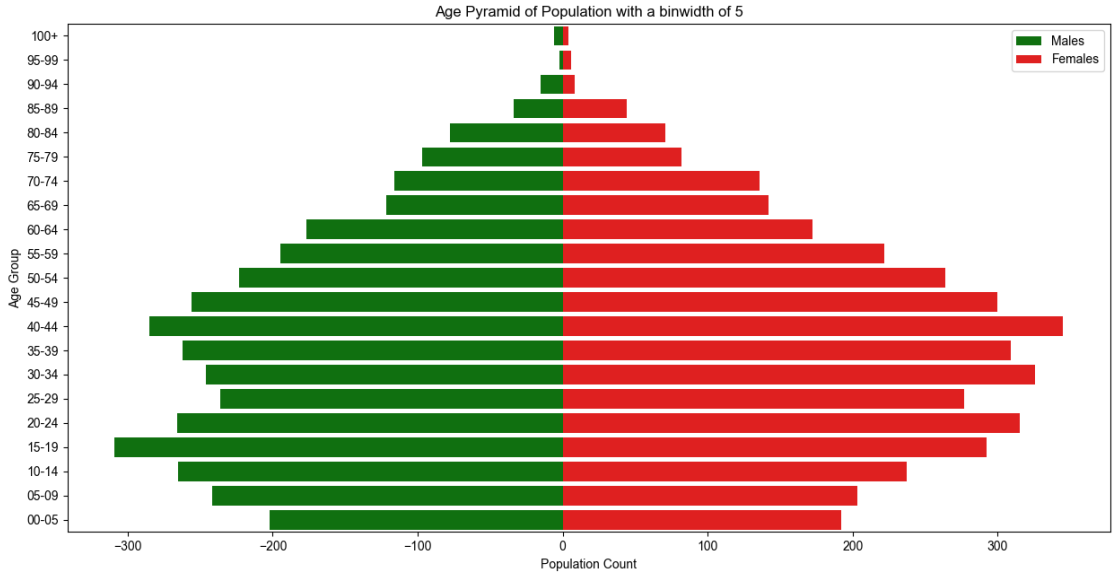


*Figure 5 Picture of Age pyramid of the town population with a binwidth of 5.*

The understanding of this pyramid chart shows a constrictive population pyramid, there are not too many children born into the town compared to the other age groups in the town. The shrinking on the male side and the stagnancy on the female side for ages between 19 and 30 years could most likely be as a result of students leaving the town for university education, job opportunities, or pursing and apprentice field else were. As we already know, this is a moderately sized town sandwiched between two cities and there will always be the urge for the younger demography to explore the cities for better opportunities. However, the shrinking at the latter stages of the chart from about 40 years indicates that there is quite a number of deaths in the older demography.

Calculating the crude birth and death rates I used the formular shown below; (DOH Online PDF)

i. Crude Birth Rate per Hundred Thousand = (Total number of births/Total population) X 100,000
ii. Crude Death Rate per Hundred Thousand = (Total number of deaths/Total population) X 100,000

After my calculation, the crude birth rate for the town per hundred thousand amounted to 910.17 while the crude death rate was 1337.55.

I then went on to consider possible immigrants and emigrants so as to expatiate on my previous assessment. This is to help me decide if the town is growing or shrinking. For possible immigrants I considered employed lodgers in the town because by having a job they are more likely to settle in the town while for emigrants I considered unemployed lodgers because they are more likely to leave the town

if they get employment elsewhere. Calculating the number of immigrants per hundred thousand and the number of emigrants per hundred thousand. I can say from my analysis that the town appears to be growing in population as shown below;

```
The birth rate for the town per hundred thousand is : 910.17
The toatl death rate  per hundred thousand is : 1337.55
The ratio of immigrants per hundred thousand is: 3033.9
The ratio of emmigrants per hundred thousand is: 263.82

The population analysis per hundred thousand being 2342.7 shows that the town seem to be on an increase; as the result of
our analysis is postive and you have people coming into the town to settle. This is based on our presumption that employe
d lodgers are most likely to settle in the town
```

*Figure 6 Picture showing result of my analysis of birth, death, immigrants and emigrants' rates.*

Overall, the town still retains a high demography of ages 19 – 55 with the mode age being 19 years, while the mean age is 37 years. Going by the above I can say that there is huge potential in the town with regards to the available workforce.

```
# Calculating the mode,mean and median age in the age series.

mode_of_age_series = cleaned_age_data["Age"].mode()
mean_of_age_series = cleaned_age_data["Age"].mean()
median_of_age_series = cleaned_age_data["Age"].median()
print(f"The mode of the age data series is : {mode_of_age_series[0]} years \n median of age data series: {int(round(median
```

```
The mode of the age data series is : 19 years
 median of age data series: 36 years
 mean of age data series: 37 years
```

*Figure 7 Picture of python code and printed result for mode, mean and median ages for the age series.*

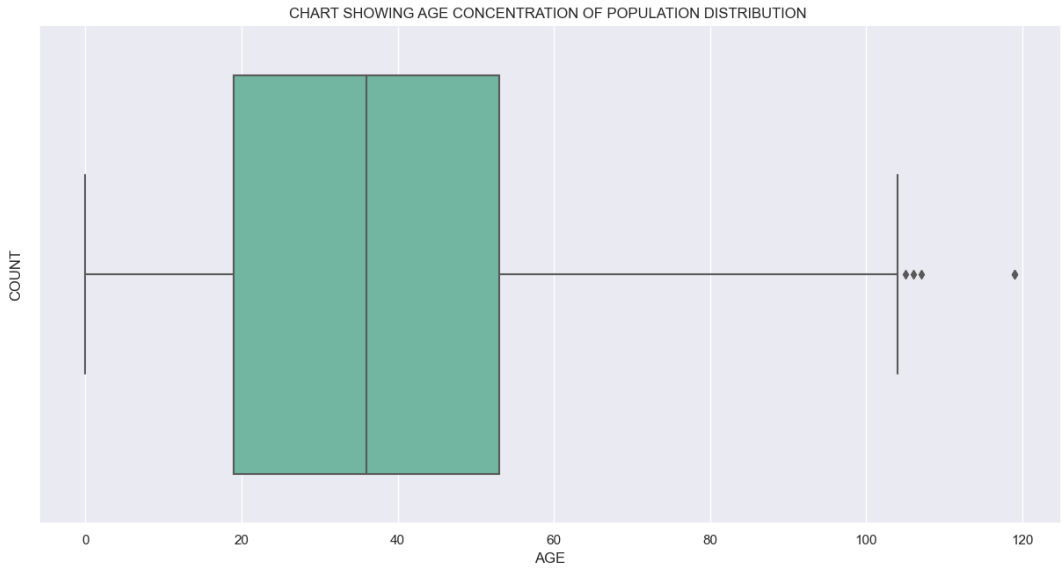


*Figure 8 Picture showing age concentration of the town population.*

## Religion

For insight into the religion series, I produced a chart showing each religion and number of persons who have preference for them as shown below;

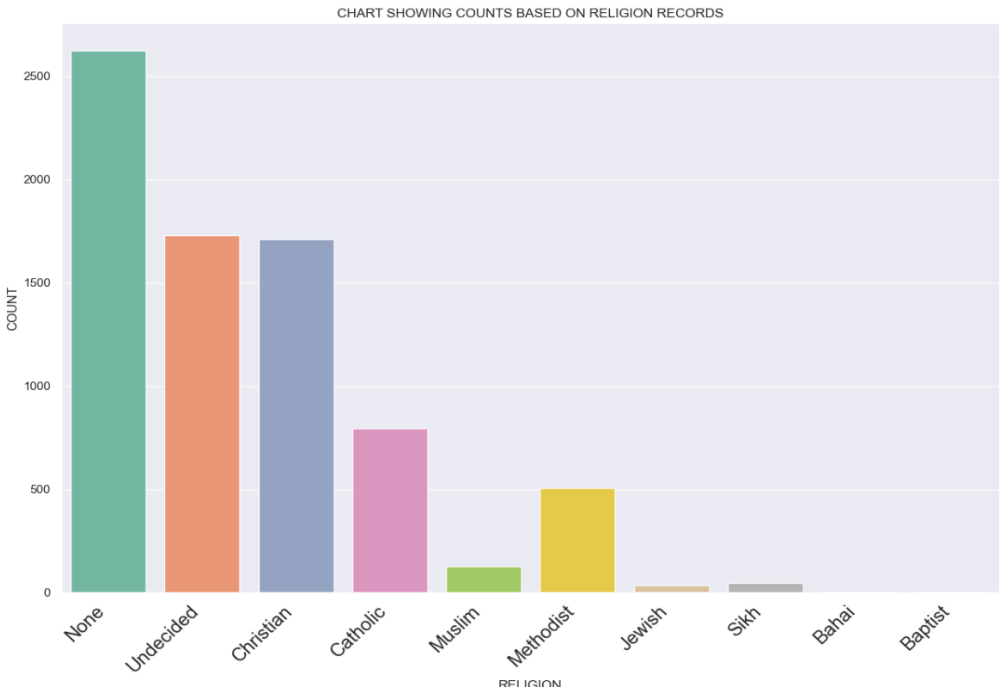


*Figure 9 Picture showing graphical representation of population for each religion*

From the above we can clearly see that the Christians would require a building for worship going by their population as the Catholics have one already. But a further analysis shows that the Christian religion have a dominant age group of between 35 to 65 years and a mean age of 55 years as we can see in the chart below. The Christian religion seem to be on a decline and need for a Christian building for worship does not seem to be of high priority going by statistics. For the undecided they are yet to make a decision on their religious preference, so we cannot say categorically what they would decide all we can is to wait for the next census.

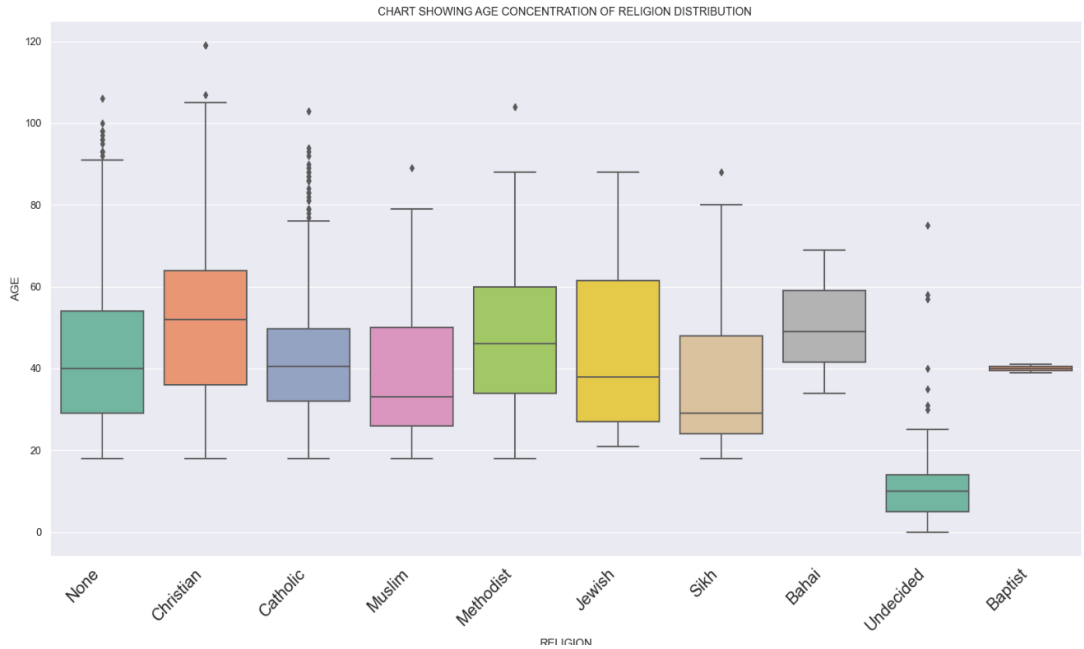


*Figure 10 Picture showing age concentration for each religion.*

## Employment

Exploring information on employment I discovered that 1/10$^{th}$ of the town are unemployed totaling 535 persons and the age concentration for the unemployed is between 35 and 55 years who are supposed to be part of the town's workforce. Going by these statistics I would strongly recommend to re-train people for new skills so as to reduce unemployment.

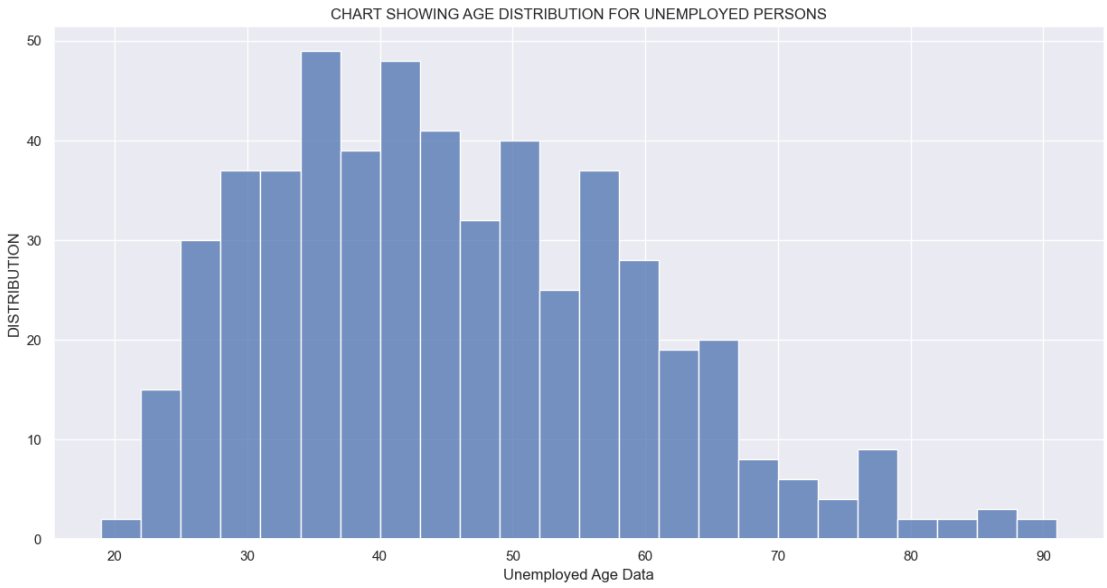


*Figure 11 Picture showing unemployment demography.*

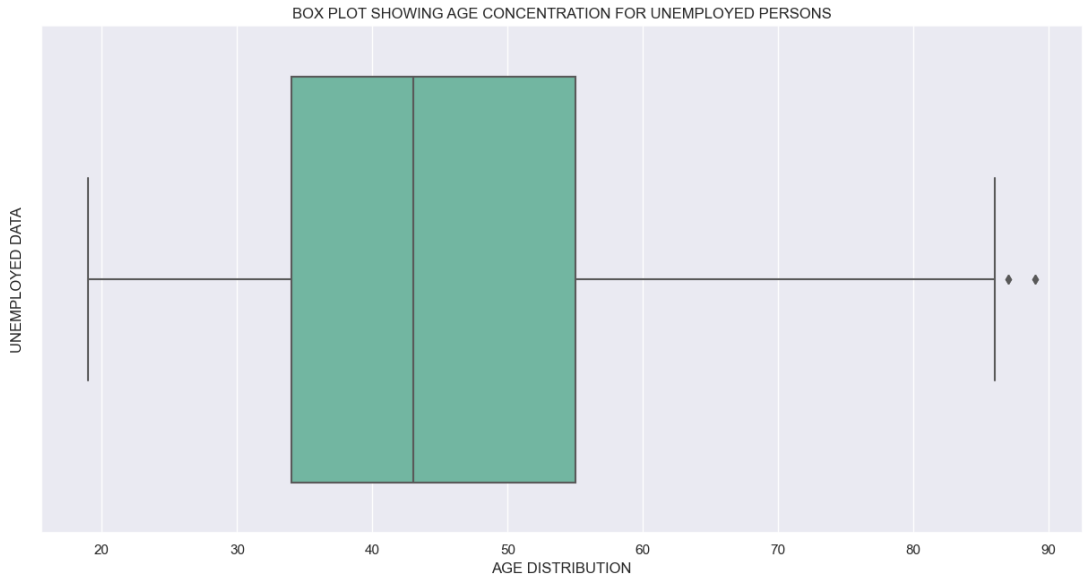


*Figure 12 Picture showing age concentration for unemployment.*

## Marriage and Divorce

According to the marriage act of 1929, in response to a campaign by the National Union of Societies for Equal Citizenship, Parliament raised the age limit to 16 for both sexes in the Ages of Marriage Act and this is still the minimum age as of today. Though parental or guardian consent is required for the marriage of a person under the age of 21 to be valid (UK Parliament, 2021).

Analyses on the marriage to divorce rates as show in figure 13 below across eligible age groups shows that persons of the younger demography have very high marriage to divorce rates of over 50% in some age groups and this trend crashes down as age increases. I would recommend counselling services for married couples especially persons between 18 – 35 years. This might be one of the causes of reduced birth as this age group is supposed to contribute significantly to the birth rate.

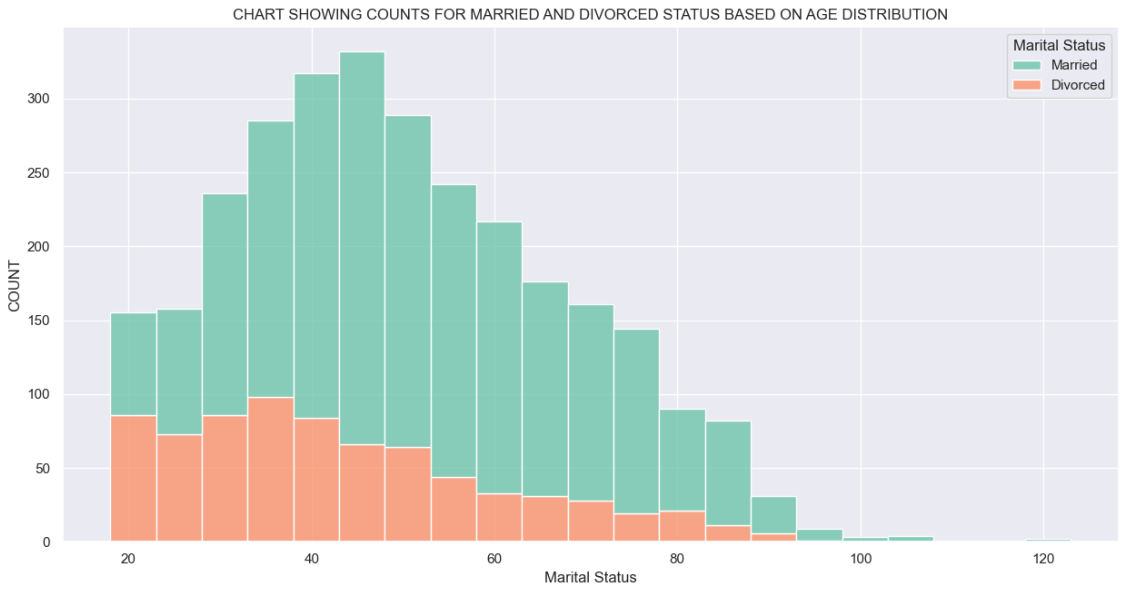


*Figure 13 Picture showing how number of married compares to number of divorced across all age groups.*

## Households

Exploring the data to get the number of possible households. I selected all "head" under the relationship to head of house column. Counting them I discovered that the number of households are 2,879 which accounts for 1/3$^{rd}$ of the town's population. This high number of households could possibly be as a result of the high divorce rate in the younger population. There is also high occupancy count in the households as the average occupancy per household is 3.75 persons per household. It appears to be that there is a shortage of available houses as majority of the households are moderately sized families with too many lodgers who are over the age to own a house thereby making a large percentage of households overcrowded. This should be taken care of by building high-density housing on the plot of land which the local government wishes to develop for the town (Shelter_England, 2021).

Number of rooms

| Number of rooms | Maximum number of people allowed |
|---|---|
| 1 | 2 |
| 2 | 3 |
| 3 | 5 |
| 4 | 7.5 |
| 5 | 10 |

Floor space of each room

| Room's floor space in square feet | Room's floor space in square metres | Maximum number of people allowed |
|---|---|---|
| 50 - 69 | 4.6 - 6.4 | 0.5 |
| 70 - 89 | 6.5 - 8.3 | 1 |
| 90 - 109 | 8.4 - 10.1 | 1.5 |
| 110 | 10.2 | 2 |

*Figure 14 Picture showing number of persons per room ratio (Shelter_England, 2021)*

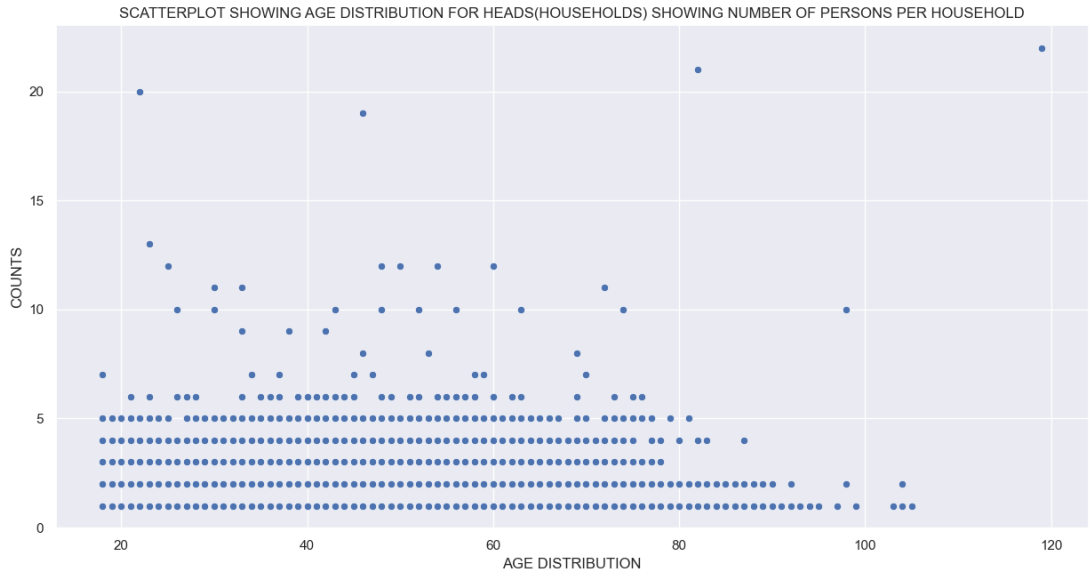


*Figure 15 Scatterplot age distribution, showing households and number of occupants*

## Commuters

For possible commuters, these categories of persons were considered as non-commuters, these include persons working in supermarkets, retail, restaurants, pubs, bars and anyone below 16 years as they were

considered to be in basic or high school which should be available in the town. After analysis I discovered that 80.78% of the town are possible commuters and by this, I would recommend a rail line from the town to nearby cities to aid commuting. Also, the rail line will provide economic value to the town in the near future.

### Elderly

For the elderly, who are persons with age 65 years and above according to Conduent Healthy Communities Institute (Conduent Healthy Communities Institute, 2021). They have a percentage of 12.7 % and also 50% of this population are likely to suffer from depression as a result of being alone as published by National Institute on Aging (NIA) whose publication states that studies have shown that loneliness and social isolation are associated with high rates of depression (NIH National Institute on Aging (NIA), 2021). I will recommend social care and activities that will encourage interaction to reduce the risk of depression in elderly persons. Persons considered likely to fall into depression were widowed, divorced and singles because they were more likely to be staying alone.

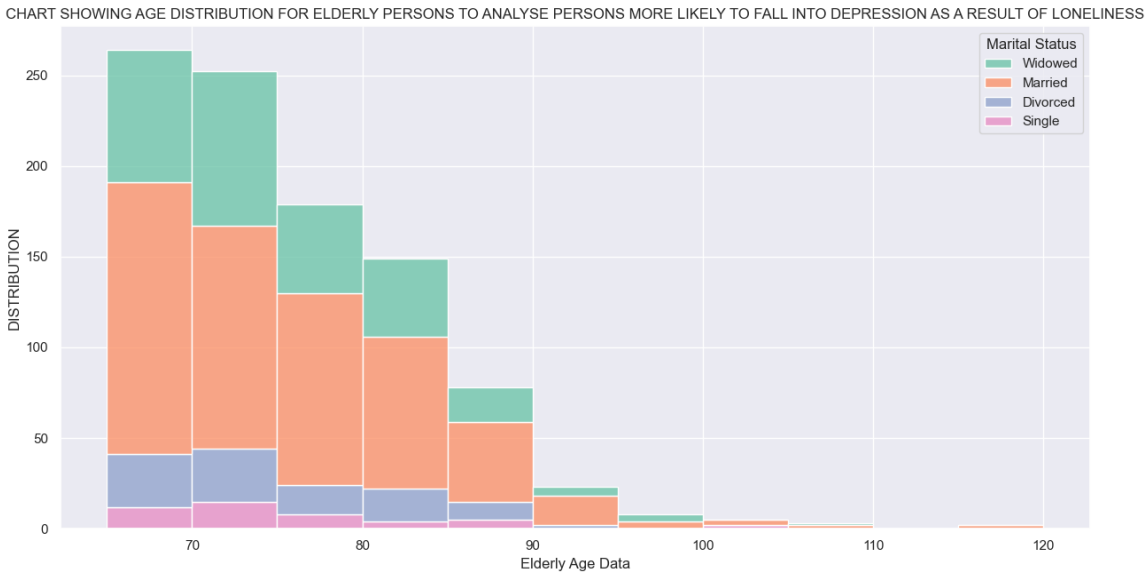


*Figure 16 Picture showing elderly persons more likely to fall into depression*

## Conclusion and Recommendation

i. Building a rail line is a necessity for the town to ease commuting.
ii. Providing social and mental care for the elderly.
iii. Building of high-density housing on unoccupied plot of land which the local government wishes to develop.
iv. Provide counselling services for marriages especially persons aged between 18 and 35 years.

# References


Conduent Healthy Communities Institute. (2021, 3 23). *How are the Different Populations Defined? - Conduent Healthy Communities Institute*. Retrieved from https://help.healthycities.org: https://help.healthycities.org/hc/en-us/articles/219556208-How-are-the-Different-Populations-Defined-

*DOH Online PDF.* (n.d.). Retrieved from www-doh.state.nj.us: https://www-doh.state.nj.us/doh-shad/view/sharedstatic/CrudeBirthRate.pdf; https://www-doh.state.nj.us/doh-shad/view/sharedstatic/CrudeDeathRate.pdf

*Jedi is not a religion, Charity Commission rules - BBC News*. (n.d.). Retrieved from BBC News: https://www.bbc.co.uk/news/uk-38368526

NIH National Institute on Aging (NIA). (2021, 7 7). *Depression and Older Adults | National Institute on Aging.* Retrieved from https://www.nia.nih.gov/: https://www.nia.nih.gov/health/depression-and-older-adults

Office for National Statistics. (n.d.). *How the Office for National Statistics is ensuring the 2021 Census will serve the public*. Retrieved from https://www.ons.gov.uk/: https://www.ons.gov.uk/census/censustransformationprogramme/census2021outputs/nationalstatisticsaccreditation/howtheofficefornationalstatisticsisensuringthe2021censuswillservethepublic

Office for National Statistics. (n.d.). *Milestones: journeying into adulthood - Office for National Statistics*. Retrieved from https://www.ons.gov.uk: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/articles/milestonesjourneyingintoadulthood/2019-02-18

Oxford Languages and Google. (n.d.). *meaning of religion*. Retrieved from OxfordLanguages: Oxford online dictionary

Shelter_England. (2021, 6 2). *Check if your home is overcrowded by law*. Retrieved from Shelter England: https://england.shelter.org.uk/housing_advice/repairs/check_if_your_home_is_overcrowded_by_law

UK Parliament. (2021). *The law of marriage - Uk Parliament*. Retrieved from https://www.parliament.uk: https://www.parliament.uk/about/living-heritage/transformingsociety/private-lives/relationships/overview/lawofmarriage-/

Wikipedia. (2021, 12 7). *Agnosticism - Wikipedia*. Retrieved from Wikipedia: https://en.wikipedia.org/wiki/Agnosticism

Wikipedia. (2021, 11 13). *Census - Wikipedia*. Retrieved from Wikipedia: https://en.wikipedia.org/wiki/Census

Wikipedia. (2021, 11 27). *Religion in the United Kingdom - Wikipedia*. Retrieved from Wikipedia: https://en.wikipedia.org/wiki/Religion_in_the_United_Kingdom