

# **CENSUS PROJECT REPORT**

## **1. INTRODUCTION**

This report is about a detailed analysis of a mock census data of a modest town between two large cities. The data was generated in a similar manner as the 1881 census of the UK.

The first part of this project contains a detailed report of the cleaning of the census data. The cleaning was properly done to correct the errors, identify wrongly imputed information, as well as the missing values and then fill them with the most appropriate values.

The following part contains the analysis carried out on the data to give the best recommendation as to (1) what should be built on an unoccupied plot of land by the local government (2) the best investment advice to give the local government. Some of the elements of the data analyzed include: age distribution, employment and unemployment trend, divorce and marriage rate, birth rate and death rate, immigration and emigration rate, population growth etc. For this project, 18 years will be taken as the start of adult age. According to the Office for National Statistics (ONS, 2018), it is not until their 18th birthday that children become adults in the eyes of the law.

## **2. DATA CLEANING**

The Jupyter notebook contains detail of how each column was cleaned.

To commence the cleaning, the data was imported using pandas. Missing values as well as duplicate values were checked using the right python function. The columns were then cleaned one after the other for proper analysis.

In the case of missing surnames, the values were imputed by observing the individuals details or that of others living in the same house. Where substantial information was not gotten, the missing values were imputed to "Undisclosed".

There were no outliers in the record for the Age column. However, the values were initially imputed as floats in the raw data which was not appropriate. For better analysis, the Age values were changed to integers.

The "Relationship with head of the house" feature had four empty strings which were correctly imputed by observing their individual details and that of their fellow housemates. Similarly, the gender column which had two empty strings was correctly imputed following the same procedure. Two individuals between aged 16 and 17 were initially imputed as "head" and did not live with any other person above 18 in the same house.

Religion had 2021 null values out of which 51 are adults while the rest are under 18(minors). The minors with null religion were assigned the religion of their head of House. The remaining null values were replaced by filling them with an even distribution of the other religion in order to reduce bias being introduced into the dataset.

One individual had “Sith” religion which was converted to ‘Sikh” as it is most likely a typographical error. “Private” religion was also converted to “Undisclosed” while ‘Nope” religion converted to “None” as they mean the same thing.

Marital status had 1891 null values that are all below 18 years old. According to (GOV.UK), people must be aged 18 and over to get married. In line with the UK marriage act, they were all converted to “NA” which means (not applicable) since it is most likely they are not married at that age. There were 3 individuals that were either married or divorced within the ages of 16 and 17. There is a possibility that these people could be lying as it is almost impossible for one to be already divorced before age 18. However, it is possible for 16 and 17year olds to marry if they have their parents’ consent. (Sigourney,2021).

After proper cleaning of the dataset, the information of the data came out having the following features.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8878 entries, 0 to 8877
Data columns (total 11 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   House Number                        8878 non-null   int64
1   Street                             8878 non-null   object
2   First Name                         8878 non-null   object
3   Surname                           8878 non-null   object
4   Age                               8878 non-null   int32
5   Relationship to Head of House      8878 non-null   object
6   Marital Status                     8878 non-null   object
7   Gender                             8878 non-null   object
8   Occupation                         8878 non-null   object
9   Infirmary                         8878 non-null   object
10  Religion                           8878 non-null   object
dtypes: int32(1), int64(1), object(9)
memory usage: 728.4+ KB
```

**Figure 1: Summary of dataset after cleaning**

Also, new features were added to help for proper analysis.

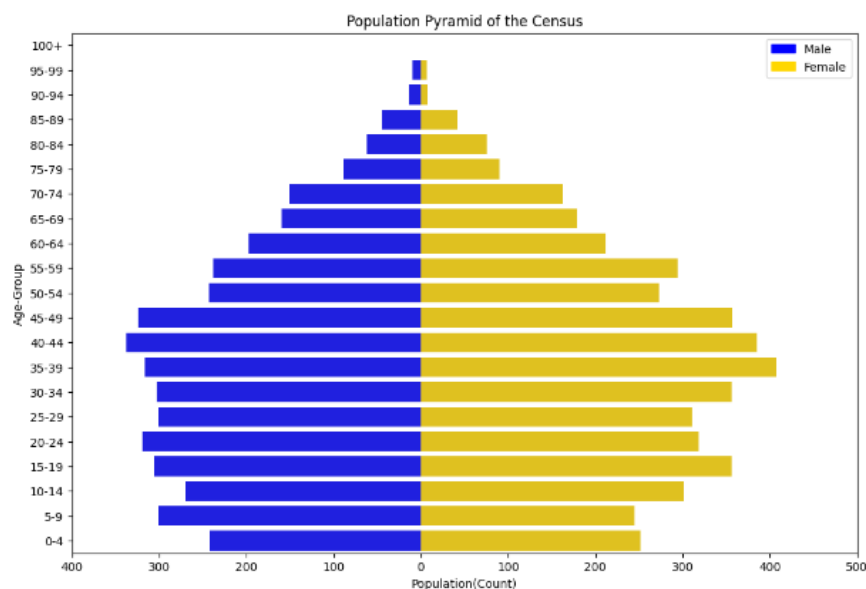
1. **Profession category:** The occupation feature had a lot of unique values which made analysis complex. To simplify this, all the people retired under different profession was classified as “Retired”. Also, both university student and PhD students were classified as University Students. For easy identification of commuters, all the occupation that is most likely to be practiced within the university environment were classified as

“University Staff”. This brought the unique values of Profession category to: Child, Student, university student, University Staff, Employed, Retired.

2. **Age range:** An age range of 5 years gap was created to help with plotting of the age pyramid and easy calculation of the death rate.
3. **Full address:** This was created to have a full address of each house for easy count of number of houses and street.
4. **Household counts:** This was created to also aid the head count of how many people staying in a particular household.

### 3. DETAILED ANALYSIS

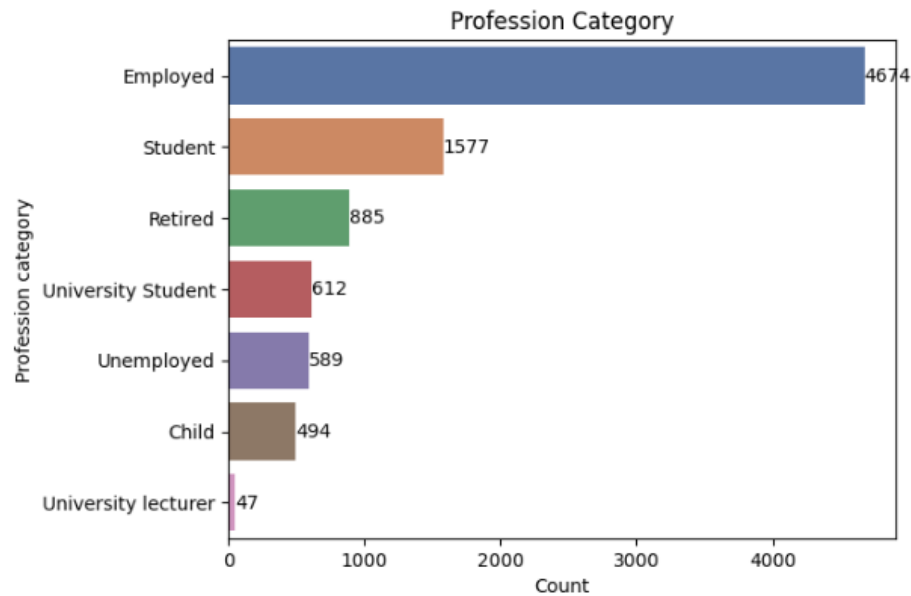
#### 3.1 AGE PYRAMID



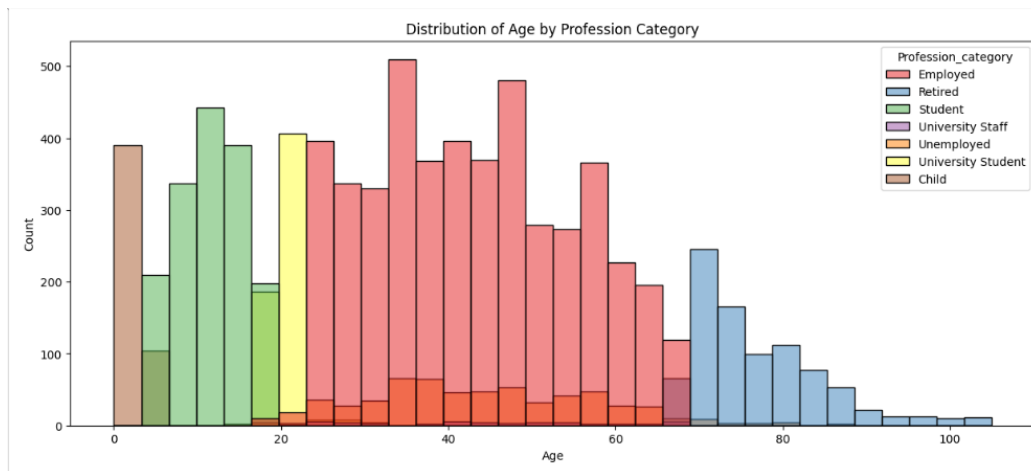
**Figure 2: Population Pyramid**

From the population pyramid above, it can be deduced that there is low birth rate. The population of the middle age is clearly higher than that of ages (0-4). The population also seemed to decline as they grow older which is understandable as old age comes with higher death rate. Also, increased population was observed around age(15-24) and (40-49)

### 3.2 EMPLOYMENT/UNEMPLOYMENT TREND



**Figure 3: Bar plot showing the distribution of the different professions**

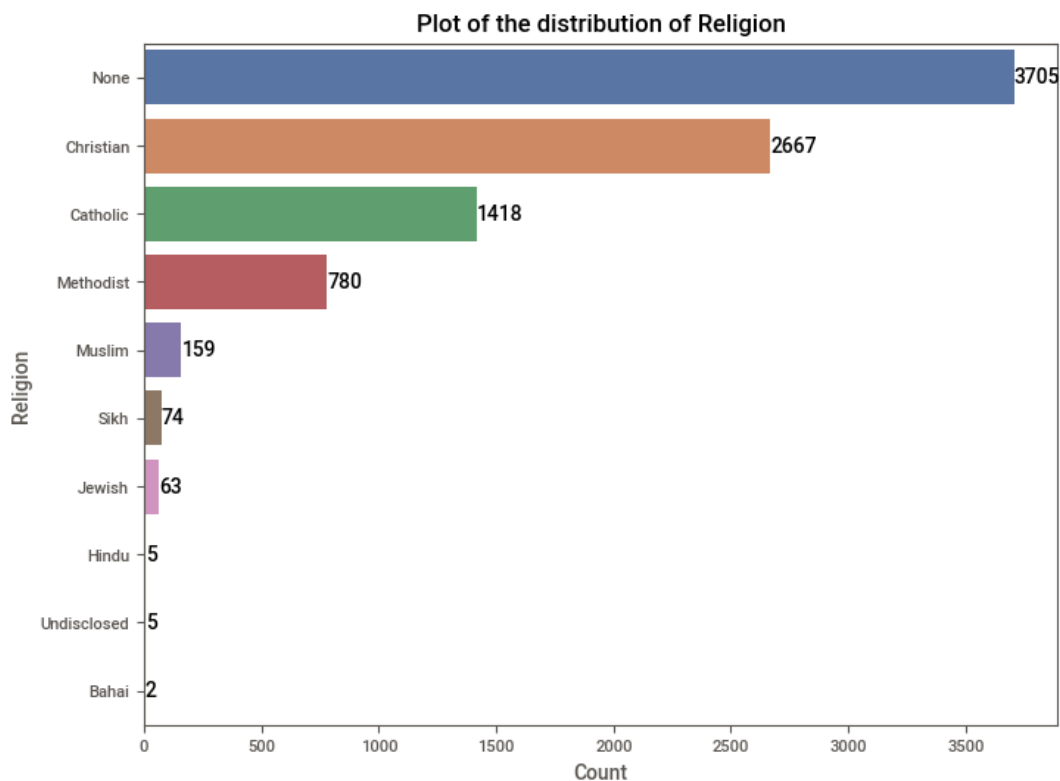


**Figure 4: Distribution of Age by the different professions**

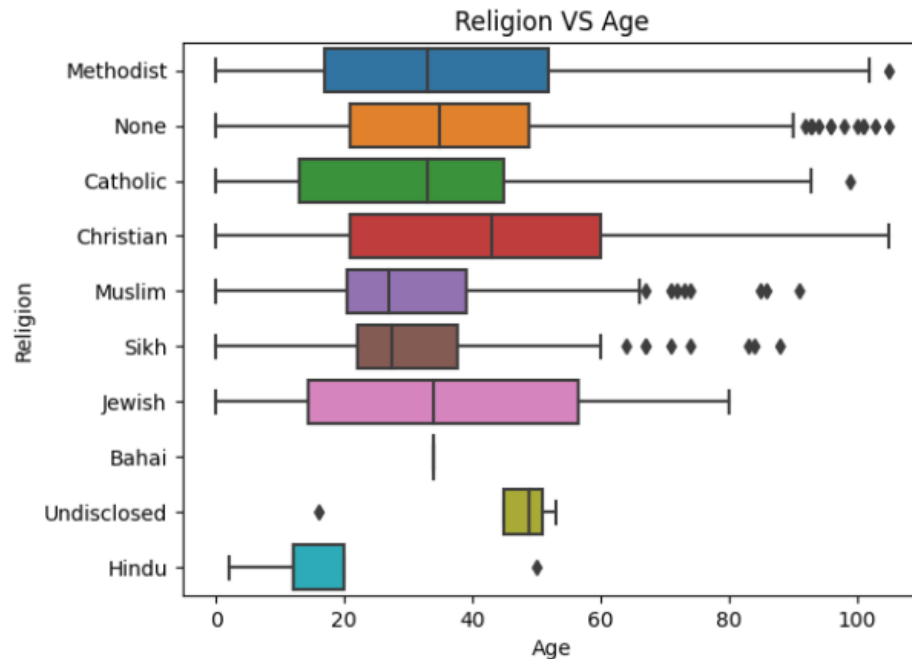
The Pensions Act 1995 introduced the equalization of State Pension age for men and women, which brought the pensionable age to 65 (GOV.UK). Since the retirement age is from 65, individuals of this age and above were placed under “Retired”. This explains the reduction in the unemployment trend towards the old age which can be seen from the plot above. Individuals below the working age (minors) and those placed under “Retired” were not considered when analyzing the unemployment trend. The unemployment rate was observed low from 18 to about 25 years because at that age majority of the people are either students or

university students. The plot above showed that a good number of the people of working age are employed while just a few of them are unemployed. Generally, the unemployment rate compared to the people of working age population is just about 10% and about 6.6% when compared to the total population. Since the unemployment trend is relatively low, training people in the town for new skills cannot be considered a priority investment. In as much as it has its benefits, there are more pressing areas that need investment in the town.

### 3.3 RELIGIOUS AFFILIATIONS



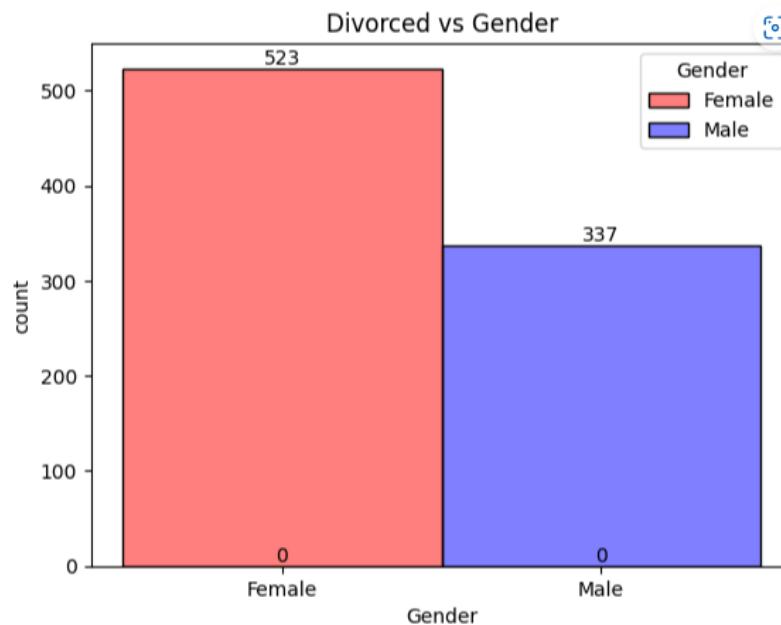
**Figure 5: Distribution of the different religion affiliations**



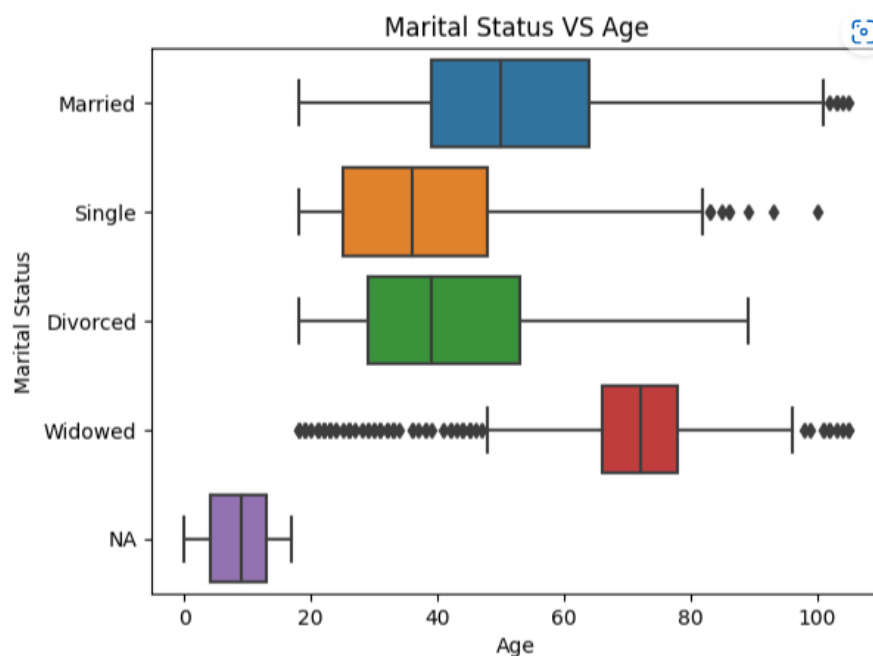
**Figure 6:Boxplot of the distribution Age by the Religion affiliations**

From the plot above, a good number of the people living in the town have no religion. Of the total population, about 33% of them have no religion. From the total number of people with religion, Christian has the highest number, although it can be said to be a combination of other Christian denominations. Catholic is the second highest religion in the town. It also has most of the people within the young age of 18 to 45 years as its members and already has a church building. From the statistics above, I do not recommend the building of another church in the town. Even though it cannot be waved off totally as a bad investment, it cannot be considered a priority investment.

### 3.4 DIVORCE AND MARRIAGE RATE



**Figure 7: Count plot of divorced population by gender**



**Figure 8: Boxplot showing the distribution of Age by Marital Status**

Looking at the plot of “Divorced VS Age”, of the 861 total divorces, 523 are female while 337 of them are male. The boxplot of Marital Status VS Age shows that divorce started happening right from young age to old age. From analysis, the number of married people was obviously higher than the divorced. Numerically, total number of divorcees was 861 which is 9.7% of the total population as against married people that are 2552 which is about 28.75% of the total

population. The divorce rate was calculated by comparing the number of divorcees with the total number of married people in the town. The resulting calculation shows that the divorce rate is 33% of the total number of married people.

### **3.5 BIRTH RATE AND DEATH RATE**

The crude birth rate and death rate was calculated using the formula below:

**Crude birth rate**

=

**Number of live births x 1000 / Estimated population size at midyear**

**Crude death rate**

=

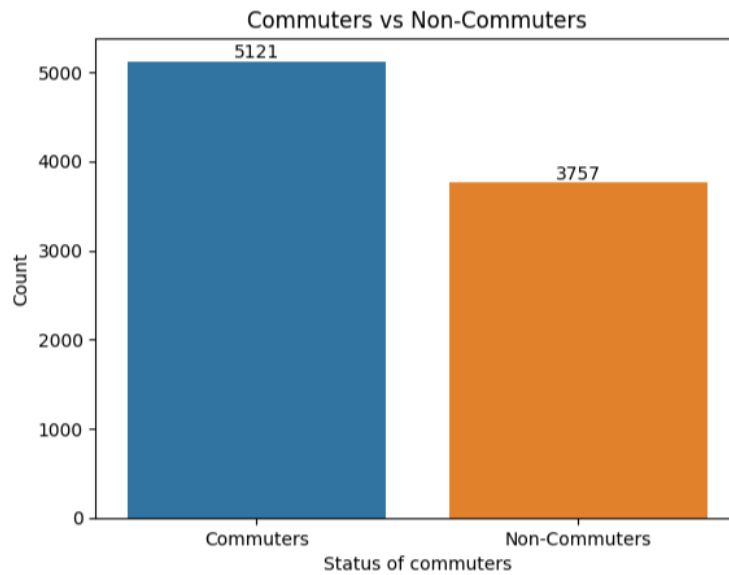
**Number of deaths / Estimated midyear population \* 1000 (University of Pittsburgh)**

From the calculation, it was deduced that the population is not growing. The current birth rate is about 10 births per thousand populations while in the last 5 years, the birth rate was about 11 births per thousand populations. To get the estimated total population for the past years, the total birth for the upcoming years was subtracted from the total population. Comparing the crude birth rate with the evolving birth rate, the birthrate fell short of 1 child per thousand populations.

For the death rate, the age range starting from the retirement age of 65 to the maximum age range was used for this calculation. 65 years was used because it is assumed that more people tend to die as they grow older. Since the number of deaths was not given, this was gotten by getting the difference between two age ranges and divided by 5 to account for the age range of 5. From the calculation, the crude death rate is about 3 people per thousand population.



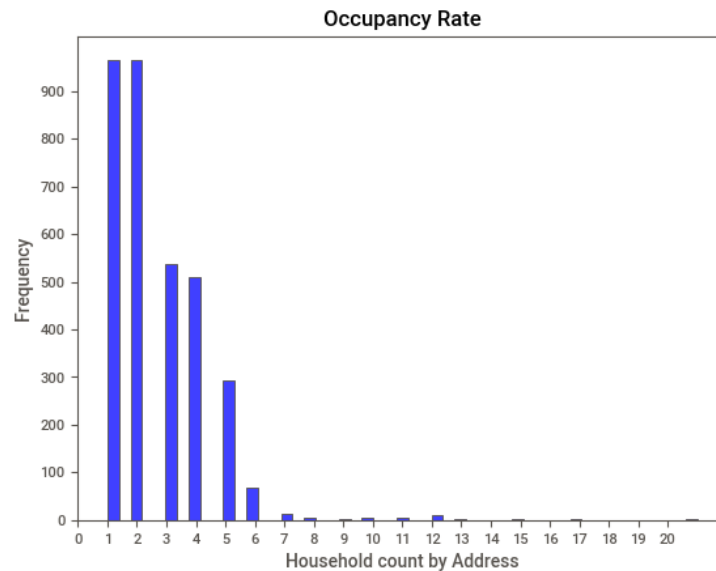
### 3.6 COMMUTERS



**Figure 9: Bar plot of Commuters and Non-commuters**

From the information given, the town does not have a university. This information guided the decision to assume commuters will be majorly university students, PhD students and university workers. Also, the employed category was critically analyzed. Since the town is a small one, certain employment categories like secondary school teachers, midwives, nurses, baristas, and community services etc were regarded as non-commuters while others regarded as commuters. Using the following analysis, about 57.47% of the population was estimated to be commuters while of the employed categories that are commuters was 95%.

### 3.7 OCCUPANCY RATE



**Figure 10: A plot showing occupancy rate**

From the plot above, it is obvious that the mode occupancy rate is 4. This means that most houses have about 4 people living in them. The median occupancy rate however is 2. This will be used to determine over occupancy. Using this means that about 1454 houses are over occupied that is about 43% of the total houses. If the median household count is increased to 3, it will mean that about 916 houses are over occupied that is about 27% of the total houses in the town. The visualization also showed that there are houses accommodating more than 10 people even up till 21 people. This can be attributed to a couple of reasons. Some of them are lodgers, visitors or divorces while others could be block of flats accommodating large families. An example is the case of “32 James Station” that has about 21 people living in it and all of them are related.

### 3.8 MIGRATION

From the age pyramid, population growth was seen between the age ranges of 15-24. These are mostly University Students. Between age range of 25 to 39, the population reduced. This could mean that most of the migrants were students who tend to leave after school while their houses are re-occupied by incoming students.

For this analysis, the total number of immigrants was calculated using the total number of individuals whose relationship with head of house is either “Lodger”, “Visitor”, or “None”. They are assumed to be people who migrated to the town. By this method the total number of immigrants summed up to 665 of the total population.

The number of emigrants was calculated by critically comparing the male divorcees and female divorcees. The female divorcees were a total of 523 of the population while male divorces were a total of 337. It can therefore be assumed that some of the male divorcees probably relocated to another city. Using the above method, the total number of emigrants amounted to 186.

The migration rate was calculated using the formula below:

$$((\text{Immigrants} - \text{Emigrants}) / \text{Total population}) * 1000$$

This brought the migration rate of the population to 54.07 per thousand populations. Based on the current census data of the number of births, number of immigrants, number of emigrants, total number of deaths and the total population, the population growth rate is about 0.68 %.

#### **4. RECOMMENDATIONS**

From the analysis, there is neither any need to build another church nor to invest in emergency medical building. This is because the birth rate is decreasing, and number of sick people is quite small. An emergency medical building might not benefit a good number of the population.

On the housing, there is a possibility that more houses will be available as people die and the migration rate is quite low now, so building houses for now might not benefit the town but in the future, it might be needed.

However, there is a need to build train station as this will benefit the University students, staffs and employees as they make up a larger number of the total population. Other occupants of the town that might not be commuting on a regular basis will also benefit from the train station as they might have the need to get to other towns and will definitely need a faster means of transportation. Also, building a train station will open the town up to more investments from foreign investors as means of transportation will no longer be much of a problem.

Also from the data, the unemployment rate is quite low, thus investing in training of the population cannot be considered a priority investment. The birth rate is reducing and the number of immigrants is not quite high with the low population growth rate, therefore investing in general infrastructure and increased spending for schooling cannot also be considered a priority investment.

However, looking at the middle-aged group which makes up the highest population, the high rate of employment at the moment with low infirmity rate signifies people will grow older and there will definitely be the need to make available the needed resources to take care of them at their old age. Therefore, investing in old age care should be the best investment option.

## 5. REFERENCES

Sigourney Lee-Smith (2021) *Legal Age for Marriage to be raised to 18 in United Kingdom*

University of Pittsburgh (No Date)

Available online: <https://hsls.libguides.com/health-data-sources>

Office for National Statistics (2021) *Milestones: journeying into adulthood*

Available online:

<https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/articles/milestonesjourneyinginto adulthood/2019-02-18>

National Society for the Prevention of Cruelty to Children (2020) *Moving Out*

Available online: <https://www.nspcc.org.uk/keeping-children-safe/in-the-home/moving-out/>  
[Accessed 20/11/2020]

The sun (2017) *In a prison far, far away 'Sith Lord' lag whinges he cannot practice Star Wars religion inside as he is banned from wearing a hoodie*

Available online: <https://www.thesun.co.uk/news/2511556/sith-lord-lag-whinges-he-cannot-practice-star-wars-religion-inside-as-he-is-banned-from-wearing-a-hoodie/#:~:text=The%20Commission%20ruled%20that%20Jediism,to%20be%20considered%20a%20religion%22.&text=In%20a%20letter%20from%20his,officially%20in%20the%20UK%20census.>

Gov.uk (March, 2023) *State Pension age Review 2023*

Available Online: <https://www.gov.uk/government/publications/state-pension-age-review-2023-government-report/state-pension-age-review-2023>