# **ABSTRACT**

This report is based on actions taken towards the data cleaning and analysis for a mock census project, the census data is very similar to a real census data, and it was generated randomly via Jupiter notebook. The data generated has been cleaned, the steps taken to clean the data has been provided in this report. Data visualization and some major statistical analysis was also undertaken to be able to offer the best recommendations on what the local government can do with an unused plot of land and/or what they can invest in based on the immediate or possible future need of the town.

**Contents**

[**ABSTRACT** 2](#_Toc89973046)

[**INTRODUCTION** 4](#_Toc89973047)

[**THE PROCESS OF CLEANING THE DATA** 4](#_Toc89973048)

[**POPULATION DEMOGRAPHICS** 6](#_Toc89973049)

[**DESCRIPTIVE ANALYSIS** 7](#_Toc89973050)

[**MOCK CENSUS PROJECT: DATA ANALYSIS AND VISUALIZATION** 8](#_Toc89973051)

[**INFIRMITY** 8](#_Toc89973052)

[**MARITAL STATUS AND DIVORCE RATE** 9](#_Toc89973053)

[**RELIGION** 10](#_Toc89973054)

[**UNEMPLOYMENT TRENDS** 10](#_Toc89973055)

[**BIRTH AND DEATH RATE** 12](#_Toc89973056)

[**MIGRATION** 12](#_Toc89973057)

[**OCCUPANCY RATE** 13](#_Toc89973058)

[**COMMUTERS** 13](#_Toc89973059)

[**RECOMMENDATION** 14](#_Toc89973060)

[**BIBLIOGRAPHY** 15](#_Toc89973061)

[Figure 1. Cleaning the data 5](#_Toc89892656)

[Figure 2. Histogram for Marital Status. 6](#_Toc89892657)

[Figure 3 Graphical Representation of an Outlier 6](#_Toc89892658)

[Figure 4 Marital Status & Relationship to head of house 7](#_Toc89892659)

[Figure 5. Age Pyramid 8](#_Toc89892660)

[Figure 6. Boxplot of Marital Status by Age 9](#_Toc89892661)

[Figure 7. Boxplot of Religion by Age 10](#_Toc89892662)

[Figure 8. Histogram Graph of the Unemployed 11](#_Toc89892663)

[Figure 9.Histogram Graph of Students 11](#_Toc89892664)

[Figure 10. Histogram Graph of University Students 11](#_Toc89892665)

# **INTRODUCTION**

Key analysis was carried out to find answers to the defined question, to this regard, recommendations were made after quality analyses were carried out to identify the demographic of the town’s population, age distribution, unemployment trends, divorce and marriage rates, occupancy rates predicted, religious affiliations and death and birth rates.

# **THE PROCESS OF CLEANING THE DATA**

All empty spaces were converted to NAN values to make it easier to investigate the empty spaces and fill the space up appropriately or remove totally.

In the ‘Street’ column of the mock census project dataset, there were NAN values discovered in the dataframe which were replaced with the street name of the denominator values seeing that it appropriately fits into the group based on the house numbering and street name arrangement.

A screenshot of a computer

Description automatically generated with medium confidence

A screenshot of a computer

Description automatically generated with medium confidence

Figure 1. Cleaning the data

The ‘Age’ column had some null values and spaces, some changes were made according to the information provided for the individuals affected in the dataset. For instance, where it says the person is a child, and there is no age value provided for the child, the individual is assumed to be a baby with the ‘0’ age value. Originally, the datatype of the age values was in strings, however, to prevent issues with data visualization it was imperative to change to ‘int’ datatype.

Marital Status had so many null values, after further investigation majority of the null values were discovered to come from Children as they are not legally allowed to be married yet without parental consent. To fix this issue, people below the age of 18 were given the NA (Not Applicable) value (BBC, 2021).

Some unique values in the Religion section like, ‘undecided’, ‘Nope’, were converted to ‘None’, as it was assumed they do not have any Religion which logically makes sense. Christianity seems to be the Religion with the highest number of people considering that Baptists, Methodists, and Catholics are Christians as well. (ONS, 2020)More than half of the people in the town have no religion.

Figure 2 Histogram for Marital Status

Text

Description automatically generatedChart, histogram

Description automatically generatedA major outlier in was discovered at the age value of ‘2021’(Check fig. 3), after a proper look, initially there were assumptions that he might be a newly born and the head of house decided to put down his year of birth but after carrying out further investigation it was discovered that he has job as a fine artist which means he is an adult thereby prompting a different assumption. It has now been assumed that this individual’s age falls between the age of 20 and 21, which propelled the replacement of the age value of ‘2021’ with age value of ‘21’.

Graphical user interface, chart

Description automatically generatedThe above graph is the graphical representation of Marital Status values after cleaning the data. From the graph it is easy to see that there are more single people in the town followed by the married group and then there is the ‘NA’ group representing individuals below the age of 18.

Figure 4 Graphical Representation of an Outlier

The graph above (Marital Status vs Age) displays the outlier in this dataset which is an age value of ‘2021’, the value was replaced with age 21, this action was taken based on the assumption that the individual is between the age range of 20 – 21 having found he is gainfully employed as a Fine Artist from the data provided.

## **Table Description automatically generated with medium confidencePOPULATION DEMOGRAPHICS**

After cleaning and replacing the values, this is what the dataset looks like. No data was dropped during the cleaning as there were logical reasons not to drop them, however, confusing data information were replaced based on the assumption that there were normal human errors during data gathering. A new column was added to the dataset, it was added to help with creating the Age pyramid graph.

# **Graphical user interface, application Description automatically generatedTable Description automatically generated with medium confidenceDESCRIPTIVE ANALYSIS**

Figure 4. Religion & Infirmity

Figure 5 Marital StGraphical user interface, table

Description automatically generated with medium confidenceatus & Relationship to head of house

A picture containing table

Description automatically generated

The descriptive analysis above shows that more than 50% of the individuals in this town are not practicing any form of religion and Christianity happens to be the next highest. Also, about 1% in the town represents the number of people with infirmity which is quit a very small amount and might not have much effect on this mock census project. For ‘Marital Status’, there are lots of singles in this town and 4.1% represents the percentage of the widowed in the group making it the lowest in the group. We can identify lodgers from the ‘Relationship to head of house’ group.

# **MOCK CENSUS PROJECT: DATA ANALYSIS AND VISUALIZATION**

The ‘Age Pyramid’ graph below shows the population distribution between male and female with the more on the male side. The population rate of young people between the ages of 0-4 is moderate, making it clear that the chances of growth in population is reasonably high that is if emigration remains low. The graph shows that the population is growing steadily into old age.

Chart, bar chart

Description automatically generated

MALE

FEMALE

Figure 6. Age Pyramid

## **INFIRMITY**

Infirmity constitutes a tiny 1 percentage in this project, therefore further analysis will not be carried out on it as this 1% has little or no effect on this census data project.

## **MARITAL STATUS AND DIVORCE RATE**

Chart, box and whisker chart

Description automatically generated

Figure 7. Boxplot of Marital Status by Age

Text

Description automatically generatedGraphical user interface, text

Description automatically generatedIt can be seen from the graph that there are more single males in this town than females with a total of 3269 of both male and female making them the highest in this group. The divorce rate is about 879 in total, which is considerable fair since there has been an increase in divorce rate of late (Independent, 2021). The widowed group started from the mid-40s down to their 90s and some few outliers that is not out of place since people die due to unforeseen circumstances. The divorce rate will be used in calculating the emigration rate later in this report.

## **RELIGION**

Chart, box and whisker chart

Description automatically generated

Figure 8. Boxplot of Religion by Age

In the religion section it can bee seen from the graph that the younger ones tend to have no religion putting into consideration that children cannot decide yet on what they want regarding religion. There are more older people identifying as Christians, that could be because Christianity is one of the oldest religions. There are more younger people of about 5442 that have not identified with any religion. There are more older people that have identified as Catholics, Baptists and Methodists and even more have identified as Christians. A total of 3711 makes up for the all the denominations of Christians including Catholics (Frassetto, 2020), Baptists and Methodists.

## 

## **UNEMPLOYMENT TRENDS**

About 6% makes up for the unemployed group in this dataset, there were some people in the retired group claiming to be unemployed, which seem out of place as the UK retirement age is 65 years old, it is better to separate the retired group from the unemployed group, even though there is no more forced age of retirement (GOV.UK, 2021), however, the age 65 will be applied for the purpose of this Census project. From the graph below it is obvious that the rate of unemployment is more present in individuals between the age range of 30 to 50, however, it starts to climb from the age of 20 and starts to drop from the age of 50.

The rate of the unemployed is relatively high, the local government could consider looking into investing in employments and training, since unemployment rate seem to be climbing, also considering that there are lots of students and university students, this means future there will be lots of qualified workers that might not be able to get jobs. However, students have the tendency to emigrate from the town since their university is in another town, some University students might prefer to settle where they studied.

Figure 9. Unemployed by AgeGraphical user interface

Description automatically generated with medium confidenceChart, histogram

Description automatically generated

Text

Description automatically generated with medium confidenceChart

Description automatically generated

Text

Description automatically generatedFigure 10.Histogram Graph of University Students

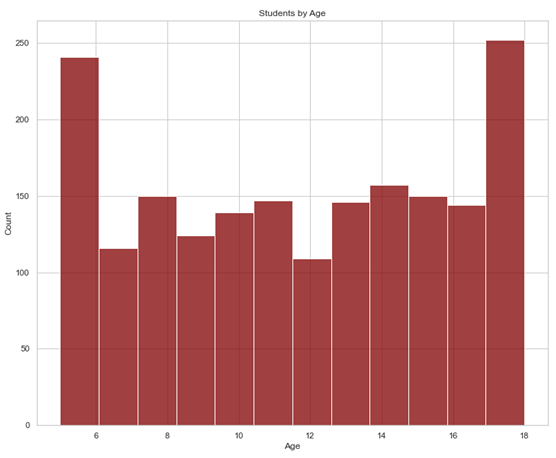


Figure 11. Histogram Graph of Student

## 

## **BIRTH AND DEATH RATE**



Birth rate was calculated by dividing Number of Births during the year by the total number of population and then multiplying by 1000. This method can be used to determine previous birth-rate based on the age of the babies.

Based on this calculation present birth rate is at 8.1 per one thousand.

Previous Birth-rate is 12.1 per one thousand.

Crude Birth Rate Change was calculated to be at 4 per one thousand

Therefore, birth-rate has dropped by 4 per one thousand

Death rate is calculated by estimating number of possible deaths during the year by using the difference age-bands for individuals over the age of 65, another method is by direction calculating the number of deaths during the year then divided by population, multiplied by 1000. Going by this formular the death rate was summed up to 11.1 per a thousand.

The crude growth rate (Birth rate – Death rate) was calculated to be at -3 meaning there was no growth in population, based on previous analysis it might seem as if there has been more emigration out of the town than deaths. (Evaluation, 2021)

## **MIGRATION**

Potential immigrants were identified by summing up the number of lodgers and visitors in the dataset while potential emigrants were derived from the number of university students and Divorcees.

Formular used:

x 1000

Population change = (CrudeBirth + Immigration\_rate) - (Deathrate + Emigration\_rate).

Going by the formular above, there are 31 immigrants coming into the town and there are about 156 emigrants moving out of the town.

Population change is equal to -128 for the mock census dataset (Evaluation, 2021).

This is informing us that population is dropping moderately by the reason of emigration

## **OCCUPANCY RATE**

A picture containing text, white

Description automatically generatedThe median of the occupancy count is at 2, which does not imply that every house in the town has only two occupiers. As can be seen on the screenshot image of occupancy detail of about 24 houses from the dataset on the left that there are several houses having more than two occupants.

Using two being the median, it produced about 1658 over occupied houses. Increasing the median to three gives about 1075 over occupied houses. About 35 houses have more than seven occupiers living in those houses which can be considered fair. This is making it clear that there is high demand for low density houses.

One can attribute high occupancy to students living together in a shared apartment or that they are large mansions full of both nuclear and extended family.

## **COMMUTERS**

Table

Description automatically generatedThere are more students commuting from the town to universities outside of the town including members of staff of these Universities outside of town since there are no Universities in the town. Some jobs that have to do with working at the outskirts of the town were also added to the commuters, like a Plant breeder, Quarry managers, Farm Managers, Hydrogeologist, and Geophysicist/field seismologist.

# **RECOMMENDATION**

From the analysis that has been carried out there seem to be a high demand for both low density and high-density housing at an almost equal proportion having observed that there are quite a lot of singles of about 3269 living in the town who might be looking at moving to a new apartment. Also, in this same dataset there are many married people with many children under the age of 18 living together. The local government can investigate investing in both low density housing and high-density housing.

6% makes up for the number of the unemployed, which is high therefore investing in employment and training might be imperative for this town. (Foundation, 2019)

Erecting a building for religious practise might not be the best option when it comes to investment by the local government. Catholics, Baptists, and Methodists are also known as Christians, however for the purpose of this mock census it is better to leave them as they are since they worship in different places. since it is highly likely that the younger ones that have not yet identified with any religion will not attend, still difficult to predict since there are children that have not identified yet with any religion in the mix.

There seem to be some considerable number of commuters in the town, adding University students including PhD students to the mix, who commute regularly and those individuals whose work base requires commuting a lot. For people in this category, constructing a train station will be of great benefit to them.

In this town there are people likely to get into their old age in the nearest future, so investing in old age care is also a welcome idea if emigration remains low.

Increase spending for schooling will benefit the children in the town seeing that there are many school age children residing in the town, children under the age of 18 make up for about 24% of the population and futuristically there will be more children based on the number of singles in their marriageable age.

# **BIBLIOGRAPHY**

BBC. (2021, November 19). *MPs back raising minimum marriage age to 18 to protect children*. Retrieved from BBC: https://www.bbc.co.uk/news/uk-politics-59344731

Evaluation, M. (2021, November 12). Retrieved from The Cohort Component Population Projection Method: https://www.measureevaluation.org/resources/training/online-courses-and-resources/non-certificate-courses-and-mini-tutorials/population-analysis-for-planners/lesson-8/lesson-8-the-cohort-component-population-projection-method

Foundation, T. H. (2019, December). *Employment and unemployment*. Retrieved from How does work affect our health: https://www.health.org.uk/news-and-comment/charts-and-infographics/unemployment

Frassetto, M. ,. (2020, November 11). *Britannica*. Retrieved from Roman Catholism: https://www.britannica.com/topic/Roman-Catholicism/Popular-Christianity-c-1000

GOV.UK. (2021, May 07). *Working after State Pension age*. Retrieved from https://www.gov.uk/working-retirement-pension-age.

Independent. (2021, May 08). *Independent News*. Retrieved from Divorce enquiries to legal firms soar by 95% in pandemic with women driving surge in interest: https://www.independent.co.uk/news/uk/home-news/divorce-inquiries-rise-pandemic-women-b1843359.html

ONS. (2020, May 09). *ONS*. Retrieved from Exploring religion in England and Wales: February 2020: https://www.ons.gov.uk/peoplepopulationandcommunity/culturalidentity/religion/articles/exploringreligioninenglandandwales/february2020