Census Project

# Introduction

‘Every ten years, the United Kingdom undertakes a census of the population, with the most recent one having been conducted in 2021. The purpose of such a census is to compare different people across the nation and to provide the government with accurate statistics of the population to enable better planning, to develop policies, and to allocate money for certain development projects.’[(*Why we have a census*, no date)](https://paperpile.com/c/efpxAK/cPD5)

# Data Preprocessing

Glancing through the data, I realized that a combination of missing and NaN values were present so I converted missing values to NaN using the’ .replace()’ method to make the data easier to clean. Now, I can see the following

| Feature | Number of Missing Values |
| --- | --- |
| First Name | 1 |
| Surname | 1 |
| Age | 2(float numbers also present) |
| Marital Status | 2059 |
| Infirmity | 8 |
| Religion | 2109 |

## First Name

Only one record has a missing first name. Since it's only Mr Taylor with a missing first name, we could assign a random name to him. I assigned him the first name of ‘Unknown’.

## Surname

When I check for null Surname values, only index 5553 has a missing surname. From closer investigation, it can be inferred that he is the son of Ms Joyce Adams and so I assigned him the surname of his family i.e. ‘Adams’.

## Age

Two records (2206 and 4317) have NaN Age values. We can say for Christopher Johnson, his age is 3 years older than his wife because according to [(Gitnux, 2023)](https://paperpile.com/c/efpxAK/nssX). The average age disparity between couples in the Uk is 3 years. For Hugh, I assigned him the average male university student age which is between 18-22 according to "[(*News and insight*, no date)](https://paperpile.com/c/efpxAK/mIJE).

Floating point numbers are also present in the Age feature. To change all age values to integers, I first converted the age values to floating points before converting them to integers by using astype('float') and astype('int') respectively.

## Marital Status

According to the [(Ministry of Justice, 2023)](https://paperpile.com/c/efpxAK/amWV), in the UK children below the age of 18 can’t be legally married. It is a crime punishable by up to 7 years in prison. So the first thing I did was to check amongst the 2059 records with NaN marital status values to check if there were any individuals above 18.

‘data[(data['Marital Status'].isna()) & (data['Age'] > 18)]’. I found out that just 3 records meet that criteria. For Mrs Hilton, upon closer investigation, I found out that she lives with her husband so I assigned her the marital status of married. The remaining two individuals that meet this criteria are university students, So I assumed that they’re single. Now that I've ensured that everyone with a marital status of NaN is below 18 and that is because they can't be legally married so I went ahead and assigned them a marital status of single.

## Infirmity

For infirmity, only 8 records have NaN values. Since there are only 8 NaN values, I replaced them with ‘unknown infection’.(The reason for this being that it’s better to wrongly classify them as having a disease and we later discover that they don’t have that ailment than the other way round).(Dr Asgari, 2023)

## Religion

There were 2109 NaN values present. The first step I took was to ascertain that all individuals with a NaN value for religion are below the age of 18. I am going under the assumption that children below 11 don’t have a religion. The remaining segment of the population that had NaN for Religion accounted for 10% of the population so I assigned them the mode value of Religion which is None.

# Exploratory Data Analysis

The total population of the given town is 8329. Out of that population, 4302 are female and 4027

are male.

For Age, the population has been grouped as follows:

* ‘The Silent Generation: >77 years old
* Baby Boomers: 59-77 years old
* Gen X: 43-58 years old
* Millennials: 27-42 years old
* Gen Z: 11-26 years old
* Gen Alpha: <11 years old’

[(Debczak, 2019)](https://paperpile.com/c/efpxAK/8ejO)

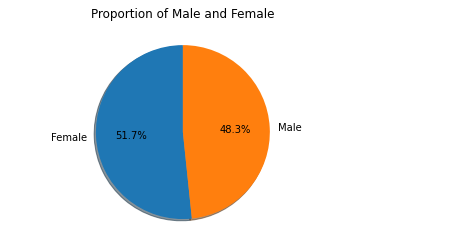


Fig 1: Proportion between male and female

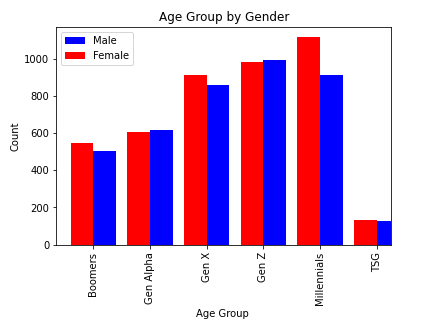


Fig 2: Population Pyramid

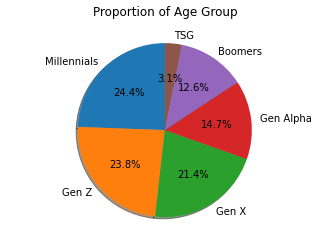


Fig 3: Age Group Proportion Across the Population

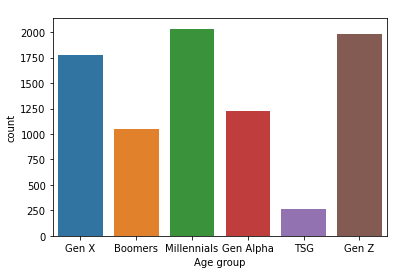


Fig 4: Age Group Count

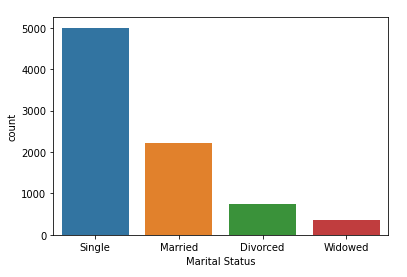


Fig 5: Marital Status Distribution Over the Population

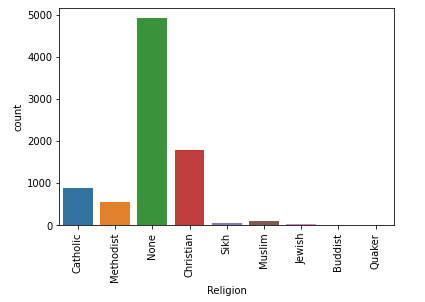
****

Fig 6: Religion Distribution Over the Population

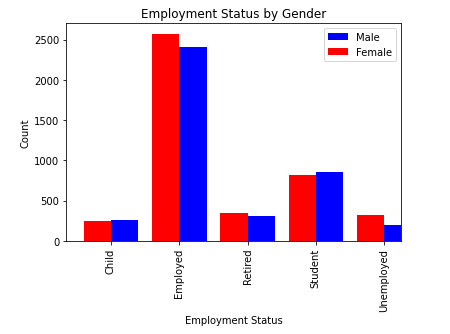
****

Fig 7: Employment Status By Gender

# Results

The general population is made up of male and female genders and from fig 1 above, it is evident that there are slightly more females than males.

From fig 2 it can be inferred that millennials(27-42 years old) make up the largest segment of the population by age group with 24.4%.

## Unemployment Rate

By using the following formula:

‘Unemployment Rate = Unemployed People / Labor Force \* 100’

The unemployment rate was determined to be **9.34%**. The labor force is a sum of both employed and unemployed individuals. [(Katara, 2019)](https://paperpile.com/c/efpxAK/M1pY)

## Child Birth Rate

Child Birth Rate was calculated using the following formula:

CBR = (Amount of children/Amount of women of Child bearing age)\*100000

Analyzing over a 10 year period, the child birth rate increased by a rate of **3053** births **per** **100,000**.

## Death Rate

The death rate over a 5 year period was gotten by using the following formula:

DR = (Difference in old population size/5)/100000.

This gave a death rate of **0.000154 per 100000 per year** over a 5 year period.

## Commuters

Students and employed individuals make up the majority of daily commuters. Commuters make up for **79.8%** of the population.

## Housing Occupancy

The formula used to derive the amount of occupants per house is shown below:

Amount of occupants = Total population/ Amount of unique households

This gives an average of **3** individuals per household. The maximum number of people in a single household is 22 and for this we can assume that the Stevenson family are a wealthy family that live in a big house and lease out some rooms because most of the occupants are members of the Stevenson family. For Toucan Divide we can assume it’s a hotel because most of its occupants are lodgers.

# Conclusion

From the insights gathered I used them to make informed decisions in answering the two questions below:

1. What should be built on an unoccupied plot of land that the local government wishes to develop?
2. Which one of the following options should be invested in?

I believe the local government should build a train station on an unoccupied piece of land because commuters make up for 79.8% of the population and that’s quite high. Building a train station would make transportation easier for students and workers who potentially need to drive to nearby major cities for work and school. Building a train station has two foreseeable benefits that I can identify, the first being taking pressure off the roads and the second is that the train station would make it easier to trade(i.e. easier transportation of bulky goods such as oil, consumer goods etc). What the second option does is that it reduces the need for lorries. [(‘What Can You Ship by Rail?’, 2018)](https://paperpile.com/c/efpxAK/7hYS)

For b, there are two possible answers, either the local government invests in old age care or they invest in schooling. However I believe the local government should invest in schooling because the population of Gen alpha and Gen Z account for 38.1 % compared to older people who would need end of life care and they account for 15.7% of the population. Also taking a look at the increase in birth rates which suggests that there will be an increase in population going by current trends.

# 

# 

# 

# References

[Debczak, M. (2019) *Here Are the Age Ranges for Millennials, Gen Z, and Gen Alpha*, *Mental Floss*. Available at:](http://paperpile.com/b/efpxAK/8ejO) <https://www.mentalfloss.com/article/609811/age-ranges-millennials-and-generation-z> [(Accessed: 27 April 2023).](http://paperpile.com/b/efpxAK/8ejO)

[Gitnux (2023) *The Most Surprising Age Difference Between Couples Statistics And Trends in 2023*, *GITNUX*. Available at:](http://paperpile.com/b/efpxAK/nssX) <https://blog.gitnux.com/age-difference-between-couples-statistics/> [(Accessed: 27 April 2023).](http://paperpile.com/b/efpxAK/nssX)

[Katara, H. (2019) *Unemployment Rate Formula*, *WallStreetMojo*. Available at:](http://paperpile.com/b/efpxAK/M1pY) <https://www.wallstreetmojo.com/unemployment-rate-formula/> [(Accessed: 27 April 2023).](http://paperpile.com/b/efpxAK/M1pY)

[Ministry of Justice (2023) *Legal age of marriage in England and Wales rises to 18*, *GOV.UK*. Available at:](http://paperpile.com/b/efpxAK/amWV) <https://www.gov.uk/government/news/legal-age-of-marriage-in-england-and-wales-rises-to-18> [(Accessed: 27 April 2023).](http://paperpile.com/b/efpxAK/amWV)

[*News and insight* (no date). Available at:](http://paperpile.com/b/efpxAK/mIJE) <https://www.hesa.ac.uk/news/17-01-2019/sb252-higher-education-student-statistics/numbers%22> [(Accessed: 27 April 2023).](http://paperpile.com/b/efpxAK/mIJE)

[‘What Can You Ship by Rail?’ (2018). Available at:](http://paperpile.com/b/efpxAK/7hYS) <http://www.up.com/up/customers/track-record/tr181120_what_can_ship.htm> [(Accessed: 27 April 2023).](http://paperpile.com/b/efpxAK/7hYS)

[*Why we have a census* (no date). Available at:](http://paperpile.com/b/efpxAK/cPD5) <https://www.ons.gov.uk/census/2011census/whywehaveacensus#:~:text=Every%20ten%20years%20the%20census,Wales%2C%20Scotland%20and%20Northern%20Ireland.> [(Accessed: 27 April 2023).](http://paperpile.com/b/efpxAK/cPD5)

Asgari, M. (2023) Fundamentals of Data Science, University of Hull Lecture.