REPORT ON THE CENSUS PROJECT

Table of Content

INTRODUCTION

In the simplest term, census is an official counting of the population which usually entails collecting of other information such as occupation,age,marital status , religion, gender, housing details and more. The information gotten is usually used by government for decision making.

In the UK, census is conducted every 10 years by the Office of National Statistics and it form the basis of determining the population (Conduct of censuses | Health Knowledge),08-12-2021.

Data Description

The data for this project was gotten from Canvas. According to the lecturer, the data was randomly generated using the Fakers package in python to emulate the format of the 1881 census of the UK where few question were asked of the population.

The town in the census is moderately sized and is in-between to larger cities connected by motorways. Though student live in the town, she has no University, and they commute to nearby city for schooling.

Objective of the Project

The project is aimed at making decision on what should be built on an unoccupied piece of land by the Local government after data cleaning and analysis. Also, what investment option should the government consider.

Data Feature

The data is made up of 11 columns with 1 numerical datatype('Age') and 10 categorical data. The total entries are 10170 and the columns include:

(1) Street Number
Street Name;
First Name of occupant;
Surname of occupant;
Age
Relationship to the
"Head" of the household
(7) Marital status
Gender
Occupation
(10) Infirmity
(11) Religion

```
<class 'pandas.core.frame.DataFrame'>                                    (2)
RangeIndex: 10171 entries, 0 to 10170                                    (3)
Data columns (total 11 columns):                                         (4)
 #   Column                        Non-Null Count  Dtype                 (5)
---  ------                        --------------  -----                 (6)
 0   House Number                  10171 non-null  object
 1   Street                        10171 non-null  object
 2   First Name                    10171 non-null  object
 3   Surname                       10171 non-null  object
 4   Age                           10171 non-null  float64
 5   Relationship to Head of House 10171 non-null  object                (8)
 6   Marital Status                7735 non-null   object                (9)
 7   Gender                        10171 non-null  object
 8   Occupation                    10171 non-null  object
 9   Infirmity                     10171 non-null  object
 10  Religion                      7673 non-null   object
dtypes: float64(1), object(10)
memory usage: 874.2+ KB
```

Summary of Numeric Data (Age)

The total number of people in the city is 10171.The average age (mean value) in the population is 35 with a Standard Deviation of 21.59. The youngest person in the population is 0 while the oldest is 105. 25% of the population are below 18years while 50% of the population are 33years or below and 75% of the entire population are 51 years or below.

Quantile statistics

Minimum	0
5-th percentile	4
Q1	18
median	35
Q3	51
95-th percentile	74
Maximum	105
Range	105
Interquartile range (IQR)	33

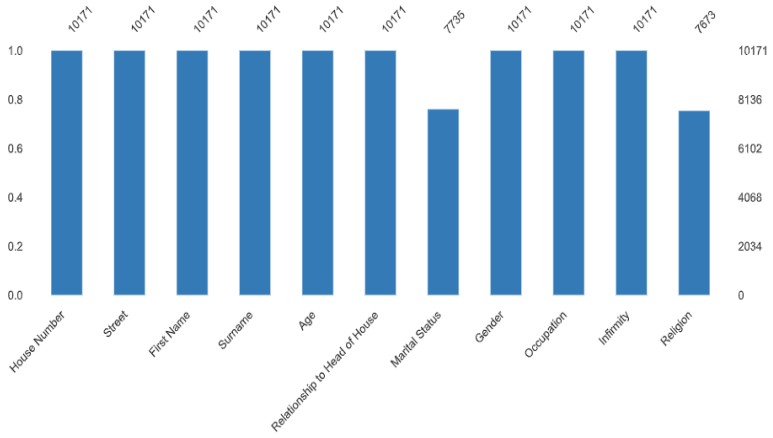
Descriptive statistics

Standard deviation	21.59718702
Coefficient of variation (CV)	0.6026919913
Kurtosis	-0.5720400975
Mean	35.8345346
Median Absolute Deviation (MAD)	16
Skewness	0.335228192
Sum	364473.0514
Variance	466.438487

Data Cleaning and Imputation

A lot of methods will be adopted in cleaning missing values, and it will be treated on a case-by-case basis. The codes can be found on the attached jupyter notebook. Missing values are found on the Marital Status and Religion columns

This chart gives a graphical outline of the culmination of the dataset. This was sorted by age. We can see that Marital Status and Religion have some values missing and its pattern shows the area of the missing value being people of a certain age range. The ‘Missingno’ library was used.



Religion, Dealing with Missing Values

Missing value in Religion for minors was replaced with their parent's because in most cases a child grows to practice the religion of their parent. Also, a child of even 12 years has already built his/her belief around the parent/guidance religion. So I imputed the minors religion to that of the previous respondent after them which is likely their parent/guidance.

Another issue with religion is error made by some respondents. There was an instance where someone inputted ‘Christ’ as their religion. This should be replaced with ‘Christian’ as there is no religion like 'Christ'. The followers of Christ are called 'Christians'.

Also, some branches of Christianity such as Baptist, Orthodoxy, Methodist, Catholic should be replace with 'Christian'. But for the sake of decision making on the piece of land (if to build a church and what denomination), I will leave Methodist and Catholic since the have high number of members.

A respondence gave his religion as 'Undecided'. This will be replaced with 'None' as it is clear that he does not believe in God but does not just want to admit it. If atheism was a religion, it would have been replaced with it. ‘Private’ will be replaced with 'None'

The blanks in this column will be replaced with 'NaN'

Cleaning Marital Status

Marital Status recognized as at 1881 were Single, Married, Divorced, Widowed. Effort are made in this section to assign any of this status to all the respondent depending on their circumstances.

Minors with missing values in their marital status were replaced with 'Single' while blanks were replaced with the mode, that is, ‘Single’

Cleaning Gender column

In one instance, ‘Female’ was mistakenly spelt as 'Fem'. This was replaced with 'Female'

The three blanks in the column were replace by ‘Female’ being the mode of the gender as this will not cause much distortion to the data or result

Cleaning Relationship to Head of House, Surname, Occupation and Others

Blanks in ‘Relationship to Head of House’ was replaced with 'Unknown' while ‘Surname’ was inferred from the details of their family member.

Blanks in ‘Occupation’ was replaced with 'Unknown'. There was an entry in Occupation spelt as 'Studen’ this was replaced with 'Student'

Every other blank in the entire data was replaced with 'None'

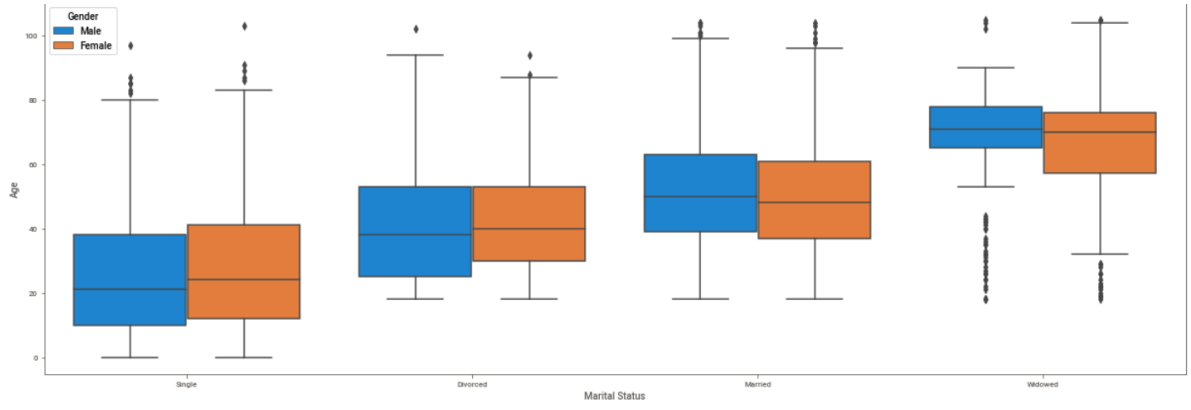
Changing float to Integer in the Age Column

The age column was changed to integer to aid data analysis.

	House Number	Street	First Name	Surname	Age	Relationship to Head of House	Marital Status	Gender	Occupation	Infirmity	Religion
0	1	Breakfast Drive	Robert	Scott	45	Head	Single	Male	Unemployed	None	Christian
1	2	Breakfast Drive	Ricky	Morris	41	Head	Divorced	Male	Youth worker	None	None
2	3	Breakfast Drive	Kathleen	Parry	26	Head	Married	Female	Accommodation manager	None	Muslim
3	3	Breakfast Drive	Steven	Parry	29	Husband	Married	Male	Administrator, education	None	None
4	4	Breakfast Drive	Barbara	Pugh	85	Head	Widowed	Female	Retired Airline pilot	None	None

Handling Outliers

Since we need to make decision based on analysing the data, outrageous values must be identified and dealt with as it can significantly affect conclusions drawn. Outliers could be as result of error during data gathering or outright lies. For the purpose of this assignment, outliers were identified by visualization using Boxplot and mathematically by using Z-score. Ways to Detect Outliers in Dataset Using Python and Pandas | Exploratory Data Analysis – DataClimbers



By using boxplot for the Marital Status, it identified some widows at a very young age as outlier especially Men. While I think that some ‘divorced young’ people provided their ‘Marital Status’ as widowed, I am also aware that it is not out of place to be widowed even at Age 18.

These were ignored but noted for further analysis. For the Age ,I set a threshold value of 3 which implies that any value presented by Z score below 3 and above 3 is identified as an outlier. People with the ages 101,102,103,104 and 105 were identified as outlier, but I had to

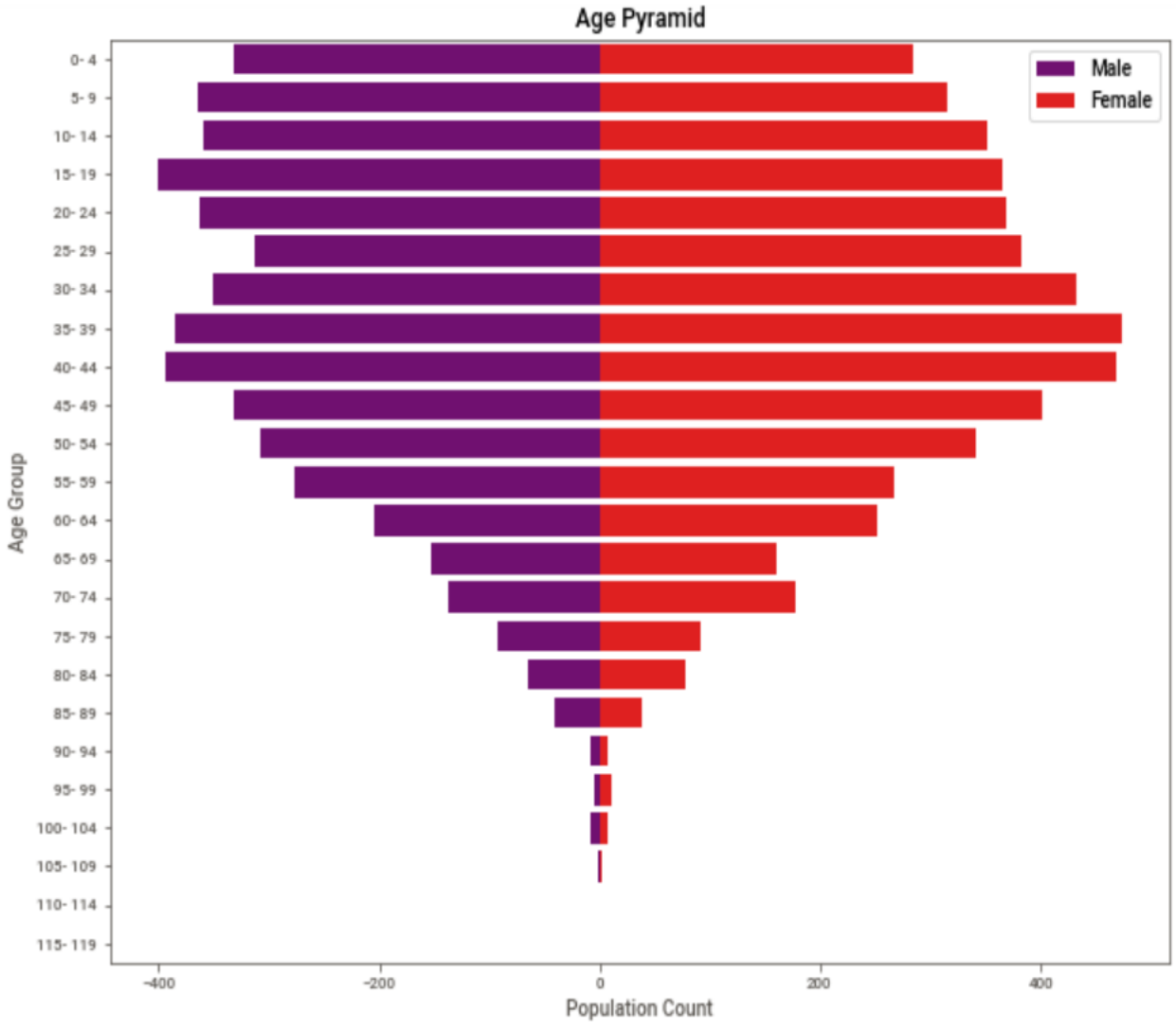
```
# using Z score to identify outliers in the Age Colunm
Mean=census['Age'].mean()
Std=census['Age'].std()
threshold = 3
outlier = []
for i in census['Age']:
    z = (i-Mean)/Std
    if z > threshold:
        outlier.append(i)
print('Outlier for the Ages are', outlier)
```

```
Outlier for the Ages are [105, 102, 104, 104, 104, 104,
```

ignore these because it can be inferred that those are their actual ages. This was done by reviewing their data such occupation (all Retired) and Relationship to Head of House.

Population Structure, Demographics and Pyramid

The makeup of a population is referred to as the population structure. It indicates the splitting of Male and Female of varied age bracket. Population pyramids is the graphical representation of the population structure which shade more light on the birth rate, death rate, population growth, dependency ratio and the life expectancy of the population.



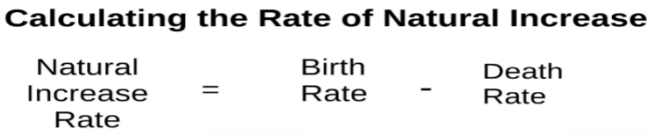
From the population pyramid of this City, she has a stable population as there is growth in 35-45 age group. This means that the crude birth and death rate is balanced. Since the birth rate and death rate is low there will be slow or no long-term natural increase.

However, the declining of this rates (birth and death) could be attributed to advanced economy, better health care system and urban migration. Also, is likely that gender equality

and women education are mainstay as this has been linked to this time of population structure. More women in the city especially of childbearing age working with increase knowledge about their bodies and access to contraception, thereby leading to decreased birth rate.

It can also be deduced that the family size of this city is highly reduced to a median of 2.6 which may be attributed to economic and social advancement.

Natural Rate



For the purpose of decision making and planning, it is important to calculate the natural rate (birth and death rate of a population) and the Net Migration(Immigration and emigration) to determine the population growth.

Subtract the birth rate from death rate and divide the difference is divided by 10.

How to calculate birth rate (Crude birth rate)

Divide total no of births in 1 year by the absolute population and then multiply the by 1000

From the analysis of population in the city, the birth rate for the current year is 10.12 while the previous year is 14.64which indicate a decline of 4.52 .The Reason for this decline could be as result of economic development, improved education for women, anxiety over future, change in social attitude, cost of bring children up, parenting less attractive or access to contraception(https://www.economicshelp.org/blog/167046/economics/reasons-for-a-falling-birth-rate/),07-12-2021

The Birth rate per 1000 for the year is 10.12
The Birth rate for the previous year per 1000 is 14.64
The difference between the previous Birth rate and the current year per 1000 is -4.52

How to calculate death rate

Divide total no of deaths in 1 year by the absolute population and then multiply the by 1000

A review of the death rate shows the rate of 11.99 which is below the world average as at the time of the census (https://www.macrotrends.net/countries/WLD/world/death-rate) ,07-12-2021

This decline could be as result of accessibility to high quality health care even by the middle class. However, there is need for the city to plan for the elderly as this will lead to higher life expectancy which implies more aged people and increase in the dependent population who may not be able to take care of themselves. Creation of care home should be one of the priorities of the decision makers.

How to calculate fertility rate

The analysis reveals that the average number of children per woman of child-bearing age in every 1000 is 274.24 . For age specific fertility rate, women between 35-40 gave more than the other age group which indicates a delay of women putting to birth until their late 30s.

Dependency Ratio

Dependency ratio shows no of people who are too young or too old to take care of them themselves (under 15 and over 65).

Calculating Dependency Ratio:

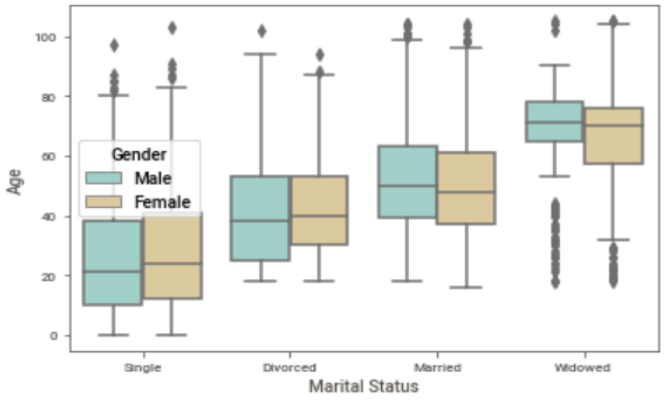
You are calculating the number of dependents (People under 15 or over 65) in a population

$$\text{Dependency Ratio} = \frac{(\%\ \text{under 15}) + (\%\ \text{over 65})}{\%\ \text{between 15 and 64}} \times 100$$

From the analysis, the number of dependents in the population is 3243 making up of 2156 minors and 1087 aged people. This is quite high and government must plan for their future.

Divorce and Marriage

From the data, both young and old ages account for the number of divorced in the town. The divorce rate is calculated is 0.41.



Total marriage is 1384.5 and it is calculated by dividing the number of married individuals by 2 as it takes two persons to make a marriage.

The number of divorced is equal to the number of female divorces which is 569

Immigration

Immigrants are people who came into city. From a review of the age group table and the corresponding population count, it reveals that the net population count of age-group 15-19 and 25-29 are almost the same. which shows that student come to the town for school and leave after graduation since the city is close to a university and their wanting to stay in the town could also mean there are student jobs or low cost of living.

Immigration was calculated by adding the lodger and visitors but excluding the divorced lodger.

Immigration rate stood at 33.52 for every 1000

Emigration

For some reasons, the number of female divorces in the city is more that the number of male divorces. This implied that male left the town after divorce. Emigrant is gotten by equating it to divorced male. The total emigrant is 375 while the emigration rate is 19.07 for every 1000

Net Migration Rate

The difference between Immigration and emigration is 14.45.

This implies that there are more people coming in than leaving the city. This will help balance the negative natural population rate (birth rate - death rate).

Population growth

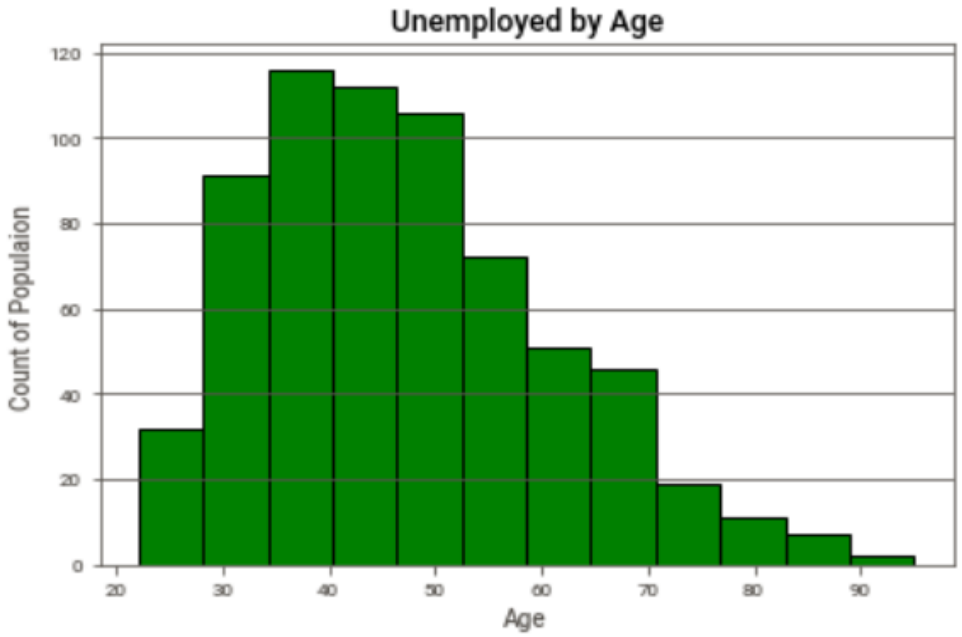
Growth rate means how the population grows during a specific time. This is influenced by factors such as emigration, immigration, Birth rate and Death rate.

The total population grow by 12.58 .This growth is directly attributed to net migration rate as there is a decline in natural due to declining fertility rate

Unemployment Rate

This rate measures the rate of people who are not employed. And it affects the real income of people and reduce the output of the economy
https://www.investopedia.com/articles/economics/10/unemployment-rate-get-real.asp,07-12-2021

Unemployment rate in the city was calculated to 9.59%.



This is a very disturbing unemployment rate and government should take action in reduction.

The age group prone to this is 30 – 50.

Total workforce is total population less dependent (aged and minors).

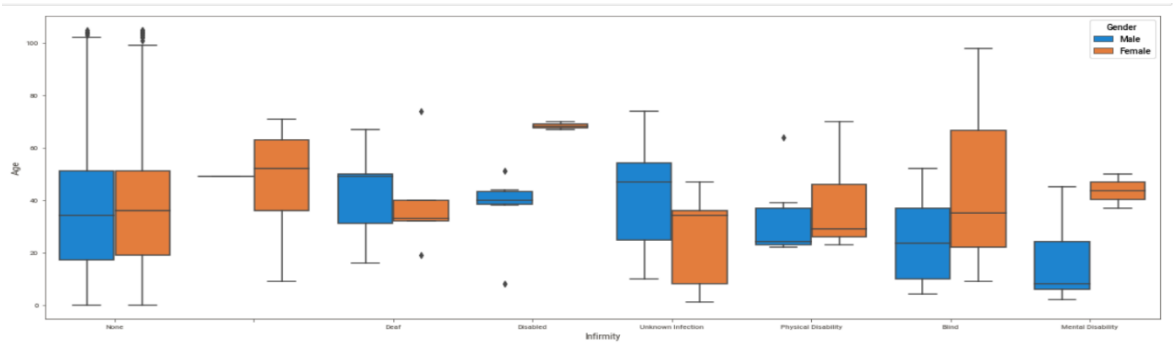
Commuters

Commuters were determined by assuming that all student in the city are commuters. Also, all employed are commuters' exception people working in jobs that does not require commuting. Occupation that contains words like community, Retail, Barista, were assumed not to be commuting.

A total of 3602 people were assumed to be commuters that is about 35.41% of the entire population.

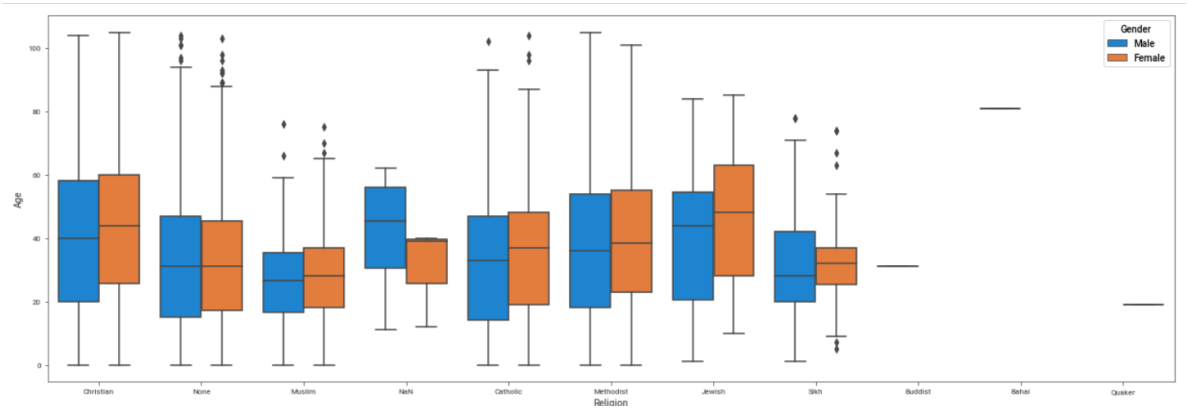
Infirmity

The town has very low infirmity rate of less than 1% indicating standard health care services



Religion

Christianity has the highest number of followers as respondents who provided ‘NoN’. This followed by Catholics. Muslim seem to be an emerging Religion in town as average age of its followers are lower than others. This may be due to immigration of Muslims into the town.



Occupancy Rates

This rate explains how many people occupies a house. From the analysis, the mean rate is 2.8 (approximately 3)

using there as a baseline, there about 1683 over-occupied houses from the total 3556. we have seen an instance where a house is occupied by over 22 people, but a further review shows an affluent family which implies that the house should be a very big one. It is likely that this kindly of situation is applicable to other houses with outrageous occupants. There is also a likelihood of hostel accommodation for student since the town is largely made up of student.

Conclusion and Recommendation

Considering that the town is made up of student (who are all commuters) and the fact there is no University in the town , I recommend that the a train station should be built with the occupied piece of land. This will not only serve the student but also other members of the workforce who have to commute to work daily. This will open the town for investment from the neighbouring cities.

There is also need to build care homes for the elderly as due to the increase life expectancy in the town as see an increase in aged people.

Also, government should urgently invest in employment and training as an unemployment rate of 9.59% is disturbing especially as it affecting the most active age-group.

Bibliography

Conduct of censuses | Health Knowledge (https://www.healthknowledge.org.uk/public-health-textbook/health-information/3a-populations/conduct-censuses [Acessed 08-12-2021]

Investopedia: Sharper insight, better investing. https://www.investopedia.com/ [08-12-2021]

Reasons for a falling birth rate - Economics Help (https://www.economicshelp.org/blog/167046/economics/reasons-for-a-falling-birth-rate/ [08-12-2021

Ways to Detect Outliers in Dataset Using Python and Pandas | Exploratory Data Analysis – DataClimbers (https://dataclimbers.com/2021/01/27/ways-to-detect-outliers-in-dataset-using-python-and-pandas-exploratory-data-analysis/ [Accessed 08-12-2021]

World Death Rate 1950-2021 | MacroTrends (https://www.macrotrends.net/countries/WLD/world/death-rate)[08-12-2021]