CENSUS PROJECT REPORT

INTRODUCTION

The aim of this report is to analyse the census 'Census 10' data of a moderately sized town with a population of approximately 7911 people and provide suggestions for upcoming services and advancements on an unutilized piece of land. The initial section of this report outlines the process of cleaning the census data, which involved rectifying data errors and filling in missing records.

The following section of this report entails important analyses that back the earlier suggestions. These analyses comprise an outline of the town's population characteristics, followed by a comprehensive evaluation of its predicted population growth, transportation users, religious beliefs, employment and occupation rates, and migration rates.

DATA CLEANING PROCESSES

The census data contains information on 11 column features of each household as presented below.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7911 entries, 0 to 7910
Data columns (total 11 columns):
 #   Column                        Non-Null Count  Dtype
---  ------                        --------------  -----
 0   House_Number                  7911 non-null   int64
 1   Street                        7911 non-null   object
 2   First_Name                    7911 non-null   object
 3   Surname                       7911 non-null   object
 4   Age                           7911 non-null   float64
 5   Relationship_to_Head_of_House 7911 non-null   object
 6   Marital_Status                5944 non-null   object
 7   Gender                        7911 non-null   object
 8   Occupation                    7911 non-null   object
 9   Infirmity                     7911 non-null   object
 10  Religion                      5895 non-null   object
dtypes: float64(1), int64(1), object(9)
memory usage: 680.0+ KB
```

Fig 1. The Census Data (Raw Data)

A detailed report of all cleaning performed on the census data can be found in the attached Jupyter notebook.

Data Cleaning

During the cleaning of 'First Name' feature, two records in row number 3689 and 4696 had blank data in the First Name column feature. These records were further investigated in their respective households, but information could not be inferred from the households. Therefore, the blank data were imputed to 'Not Given' i.e., the information were not given by the individuals during the census.

The 'Age' feature was a float data type, and it was converted into an integer data type, because the census data is of demographical interest and need not be a float data type, and there were no blank data, null or missing values in the Age column feature. However, an additional column 'Age Group' was created, it contains age entries in band of 5 years for better visualisation plots and population pyramid.

In 'Relationship to Head of House' feature the entry 'Neice' was misspelt and imputed to 'Niece', also there were no blank data, null or missing values, but three individuals (row

numbers: 4446, 4828, and 7090) were reported as ‘Head’ of households and they were of ages below 18 (they do not have the responsibility of filling the census details). Further investigations were conducted on these records by inferring information from their respective households. In one of the households, there was a female of age 16, she was reported as single and head of household, and had a child of age 0. In the second household, there was a female of age 15, she was reported as divorced and head of household. In the third household, there was a male of age 16, he was reported as head of household, married to a female of age 19, and they have a child of age 2. These three households (six records) were dropped from the census dataset, because it was evident there are two major inconsistencies or a potentially illegal relationship in these households.

During cleaning, Marital Status column feature had 1964 NaN entries, the entries were records of individuals under the age of 18, and they had Occupation as ‘Child’ and ‘Student’. Additionally, because Marriage is lawful for individuals who are 16 years old or older with parental consent (Marriage Act, 1949:s3), also additional exemption has been given to individuals who live independently, because it is possible for children under the age of 18 to leave their parents' home before they turn 18 (NSPCC, 2020), and also persons below 18 years are no longer allowed to marry or enter a civil partnership, even if they have parental consent, and it is now illegal and criminal offence to exploit vulnerable children by arranging for them to marry, under circumstances whether force is used (Marriage and Civil Partnership (Minimum Age) Act: 2022). The NaN entries were imputed to ‘Single (Minors)’ and ‘Single (Adults)’. Two individuals were widowed at age 18, these records were not dropped from the census dataset, because it is uncommon for someone to experience spousal loss at the young age of 18, however, it is plausible for someone to lose their spouse early in their marriage.

There was one blank data (row number: 7591) in Gender column feature, information was inferred from the individual’s household, and it was evident that the person involved is a daughter to the head of the household. Therefore, the blank data was imputed to ‘Female’.

During cleaning of Occupation column, there were two blank data in row numbers 418 and 2881. In general, because Occupation is a personalised entity and the information cannot be inferred from their households, the blank data in the occupation column were imputed to ‘Not Given’ i.e., the information was not given by the respondents during the census. Additionally, the Occupation column was grouped into a new column ‘Occupation Category’ which include Employed, Unemployed, Student, University Student, Child, and Retired.

There were 11 individuals having blank data in the Infirmity column, because Infirmity is a personalised entity and the information cannot be inferred from the household, the blank data were imputed to ‘Not Given’ i.e., the information were not given by the individuals during the census.

During cleaning of Religion column, there were two religions "Sith" and "Private" which looked to be deliberately false inputs. Sith, a religion which has been banned (Robocade, 2023), had 2 entries, and Private is considered to be a personal error, and had 10 entries. These entries were imputed to ‘None’. Additionally, 2009 individuals under the age of 27 had NaN as their religious status. Children and young people are free to be of any or no religion, their parents can help them make decisions around religion, but: a parent cannot force a child or young person to adopt a religion, and a parent can’t force a child or young person to stop following a religion (UNCRC, 1990: a14). The NaN entries were imputed to ‘Undeclared’.

Population Demographics

The finalised census data will have the following features after being cleaned:

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 7904 entries, 0 to 7910
Data columns (total 14 columns):
 #   Column                          Non-Null Count  Dtype
---  ------                          --------------  -----
 0   House_Number                    7904 non-null   int64
 1   Street                          7904 non-null   object
 2   First_Name                      7904 non-null   object
 3   Surname                         7904 non-null   object
 4   Age                             7904 non-null   int32
 5   Relationship_to_Head_of_House   7904 non-null   object
 6   Marital_Status                  7904 non-null   object
 7   Gender                          7904 non-null   object
 8   Occupation                      7904 non-null   object
 9   Infirmity                       7904 non-null   object
 10  Religion                        7904 non-null   object
 11  Occupation_Category             7904 non-null   object
 12  Age_Group                       7827 non-null   category
 13  Household_Occupancy             7904 non-null   int64
dtypes: category(1), int32(1), int64(2), object(10)
memory usage: 1.1+ MB
```

Fig 2. The Cleaned Census Data

To aid our analysis, these features has been included:

- **Age Group:** A column of ages in groups (5-year bands).
- **Occupation Category:** A column of grouped occupations which include Employed, Unemployed, Child, Student, University Student, and Retired.
- **Household Occupancy:** A column which denotes the sum of all members in a house.

DETAILED ANALYSIS

Demographics

The summary statistics provide information about the average age of the population, which is 35 years. The minimum age is 0 and the maximum age is 105. The household occupancy level, employment and unemployment trend, etc. will be examined in the following sections.

	House_Number	Age	Household_Occupancy
count	7904.000000	7904.000000	7904.000000
mean	27.842105	35.427885	4.191296
std	28.255771	21.222807	3.182515
min	1.000000	0.000000	1.000000
25%	8.000000	18.000000	2.000000
50%	20.000000	35.000000	4.000000
75%	37.000000	51.000000	5.000000
max	158.000000	105.000000	22.000000

Fig 3. Summary Statistics of the Census Data (Cleaned Data)

Population Pyramid (Population Distribution)

According to the population pyramid provided, there is a relatively smaller proportion of individuals in the younger age range, particularly those aged 0-4 years, indicating a reduced rate of childbirth. Moreover, the age bracket of 35-39 years exhibits the highest number of both male and female population, while there is a steady decline in population size for higher age groups until the oldest category.

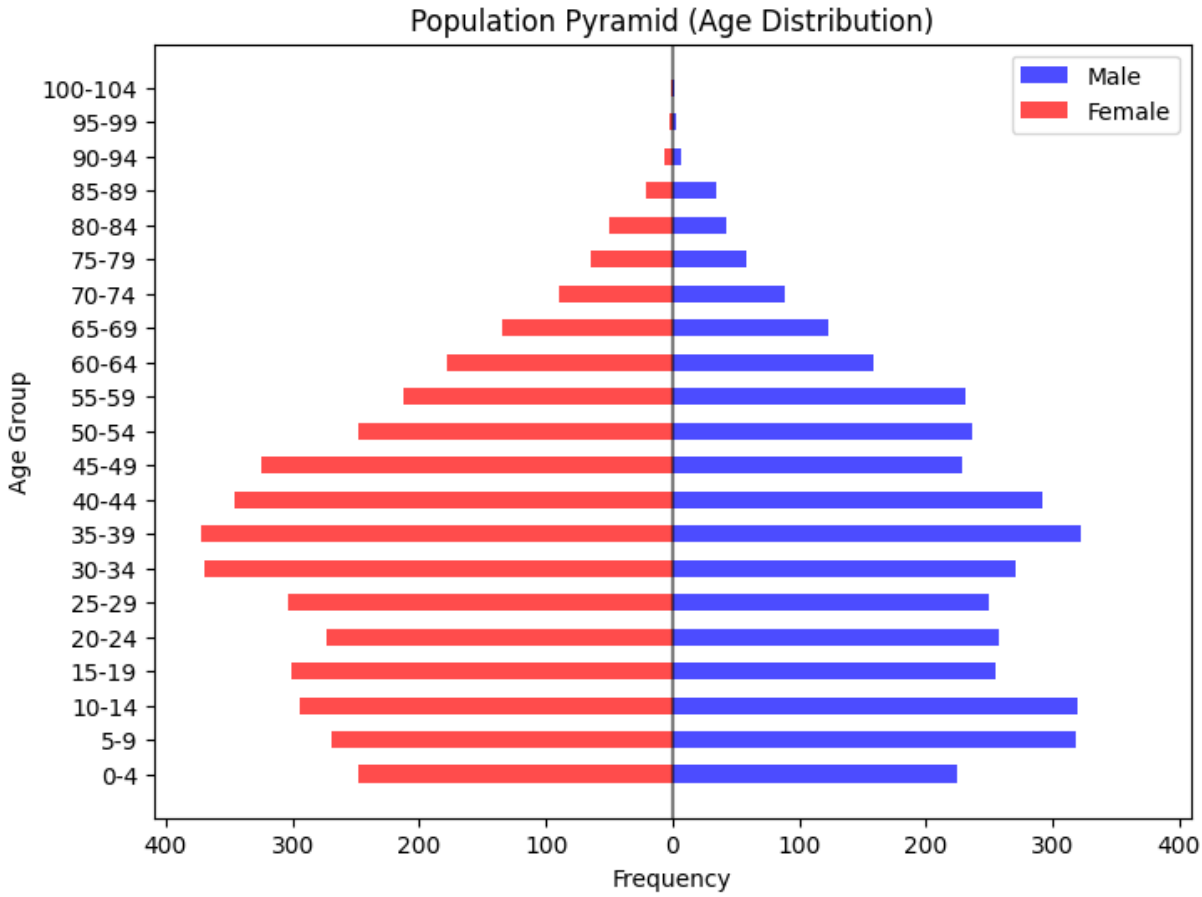


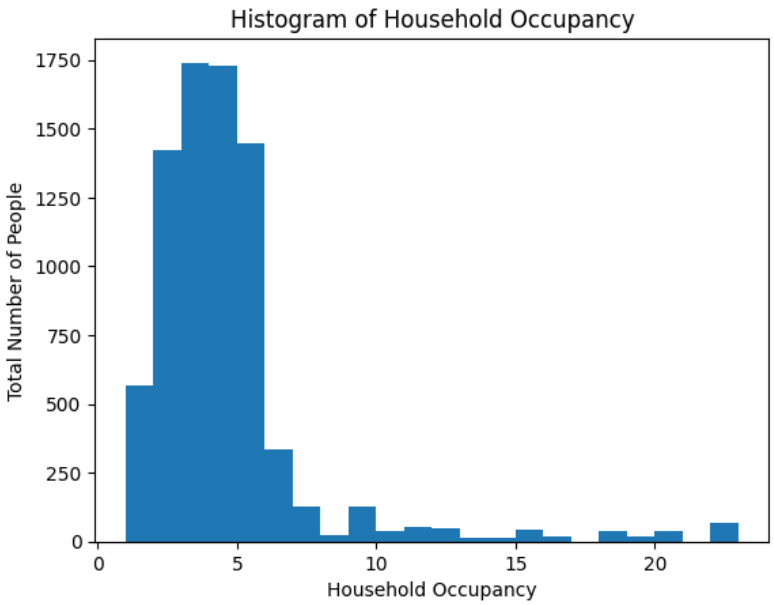
Fig 4. Population Pyramid of the Census Data

Commuter Rate

The Commuter's level was determined by obtaining the percentage of individuals in the population who are most likely to commute to nearby cities, which include occupation such as; University Student, Librarian, PhD Student, Journalist, Scientist, etc. and religious affiliation such as; Christians, Muslim, Methodist, Jewish, Sikh, Agnostic, Orthodoxy, Hindu, Bahai and Baptist, others were grouped as 'non-commuters' (individuals who are not likely to commute to nearby cities). The calculated commute rate is 39.4%, it means that 39.4% of the population commutes outside the town, while the remaining 60.6% of the population do not commute. In general, consequent to the high commuter rate in the town, it would be advisable for the government to build a train station on an unused plot of land in the town. However, this would depend on factors such as funding availability, feasibility, and demand.

House Occupancy Level

The house occupancy was determined after deriving the number of individuals living in each houses using the 'House Number' and 'Street Name' feature. The summary statistics is presented below:



Household_Occupancy	0	Number_of_Houses
1	565	565.0
2	1422	711.0
3	1740	580.0
4	1728	432.0
5	1445	289.0
6	336	56.0
7	126	18.0
8	24	3.0
9	126	14.0
10	40	4.0
11	55	5.0
12	48	4.0
13	13	1.0
14	14	1.0
15	45	3.0
16	16	1.0
18	36	2.0
19	19	1.0
20	40	2.0
22	66	3.0

Fig 5. Histogram and Summary Statistics of House Occupancy Level in the Population

Based on the summary statistics and chart provided, there are 2695 houses in the town, the median household occupancy rate of the town is 4, minimum household occupancy is 1 and the maximum household occupancy is 22, most houses in the town has household occupancy of 1 to 5 i.e., they are not fully occupied. In general, it is not advisable for the government to build high-density housing.

Religious Affiliation

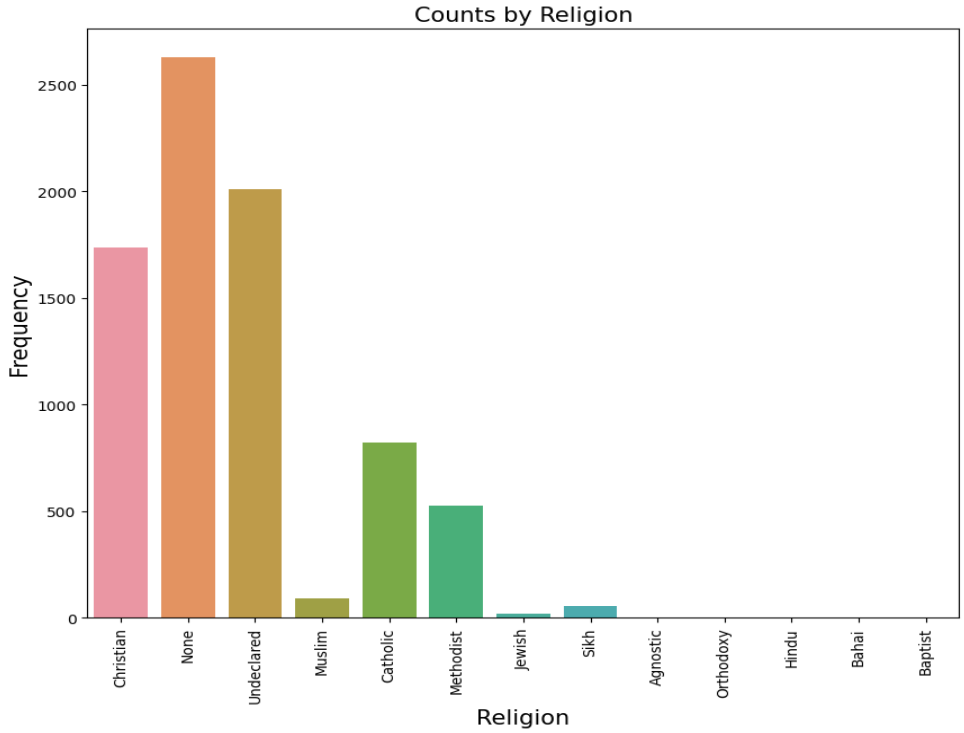


Fig 6. Count-plot of Religion Feature

According to the provided chart (count-plot) in **Fig 6**., 'None' i.e., individuals who do not belong to any religion makes up the largest group in the population, representing approximately 33% (33.26%), and Catholics make up the fourth largest group in the population, representing approximately 10% (10.39%) of the population.

Religion	count	mean	std	min	25%	50%	75%	max
Agnostic	4.0	35.750000	1.707825	34.0	34.75	35.5	36.50	38.0
Bahai	1.0	38.000000	NaN	38.0	38.00	38.0	38.00	38.0
Baptist	2.0	31.500000	2.121320	30.0	30.75	31.5	32.25	33.0
Catholic	822.0	42.991484	14.203195	18.0	34.00	41.0	50.00	96.0
Christian	1738.0	49.557537	17.259320	18.0	37.00	50.0	61.00	105.0
Hindu	3.0	43.333333	21.361960	31.0	31.00	31.0	49.50	68.0
Jewish	22.0	44.545455	15.361446	22.0	29.50	45.5	56.75	70.0
Methodist	525.0	45.036190	17.370695	18.0	30.00	44.0	58.00	89.0
Muslim	91.0	33.813187	14.100364	18.0	22.00	30.0	40.50	79.0
None	2629.0	41.807151	16.375879	18.0	29.00	39.0	52.00	101.0
Orthodoxy	1.0	42.000000	NaN	42.0	42.00	42.0	42.00	42.0
Sikh	57.0	34.754386	10.674202	19.0	28.00	33.0	37.00	70.0
Undeclared	2009.0	9.229965	5.129875	0.0	5.00	9.0	14.00	26.0

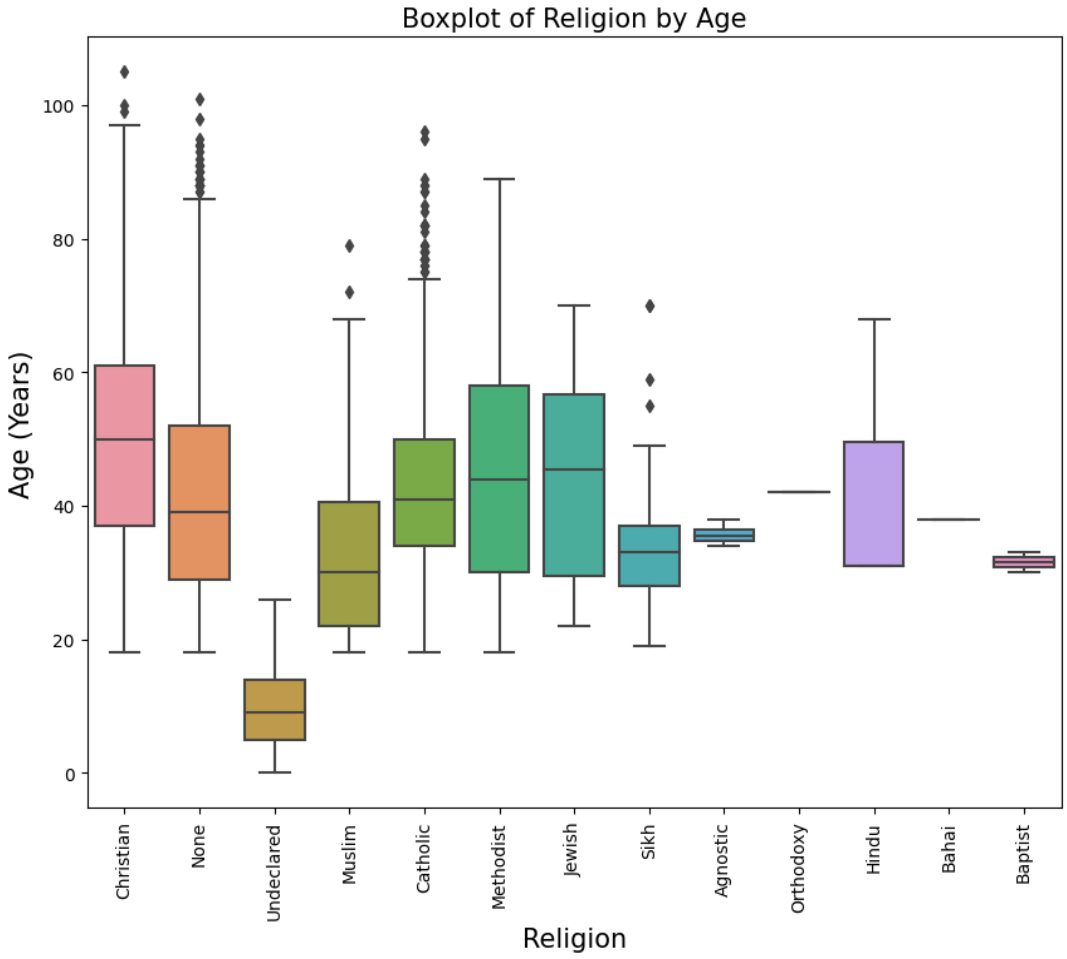


Fig 7. Boxplot of Religious Affiliations by Age

The chart provided in **Fig 7.** and summary statistics indicates that individuals of all religious affiliations are over the age of 18, except for those with undeclared religion. The individuals with undeclared religion are under 27 years old. Most people do not belong to any religious

affiliation, while only a small number of individuals belong to Agnostic, Baptist, Hindu, Orthodoxy, and Bahai religions, and they fall within the age range of 30 to 68 years. Majority of people in the town do not belong to any religion, however, majority of individuals in the town are Christians.

Infirmity Rate

The chart of the infirmity feature is provided below:

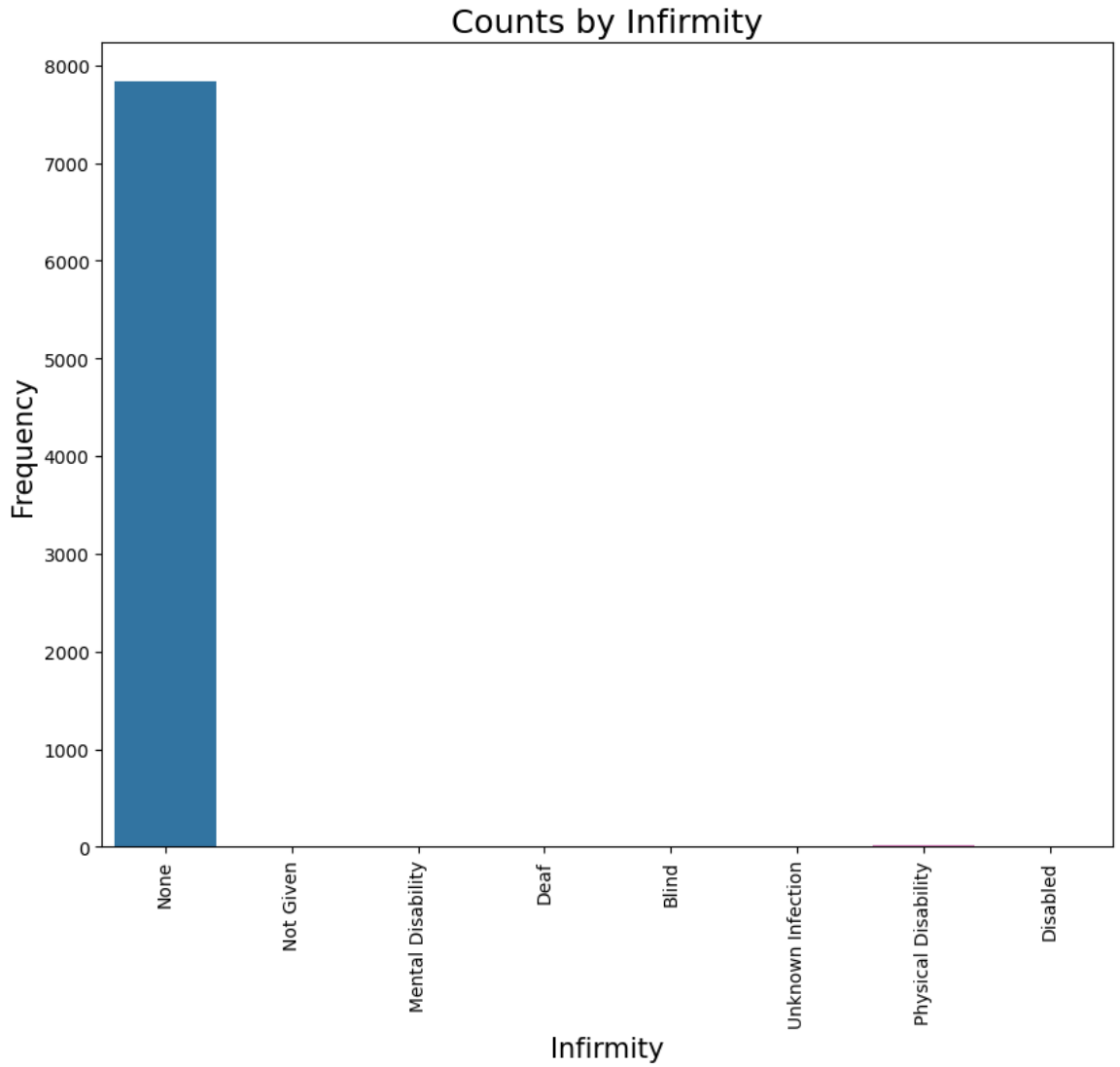


Fig 8. Count-plot of Infirmity Feature

Based on the information presented in the chart above, it appears that a vast majority of individuals in the population (approximately 99.16%) do not have any infirmity, also since there is a decline in the population of ages 0-4 as indicated by the population pyramid, and a low crude birth rate as indicated in later sections, it is not advisable for the government to build an emergency medical building in the town.

Occupation (Unemployment Trend)

The distribution of Occupation by Age in the town is presented in the boxplot below:

Occupation_Category	count	mean	std	min	25%	50%	75%	max
Child	447.0	2.100671	1.400248	0.0	1.0	2.0	3.0	4.0
Employed	4300.0	42.951395	12.108958	19.0	33.0	42.0	53.0	67.0
Retired	602.0	76.337209	6.964439	68.0	71.0	75.0	81.0	105.0
Student	1614.0	11.476456	3.915300	5.0	8.0	12.0	15.0	18.0
Unemployed	489.0	42.251534	12.532258	19.0	33.0	41.0	50.0	93.0
University Student	452.0	20.471239	1.219416	18.0	19.0	21.0	22.0	22.0

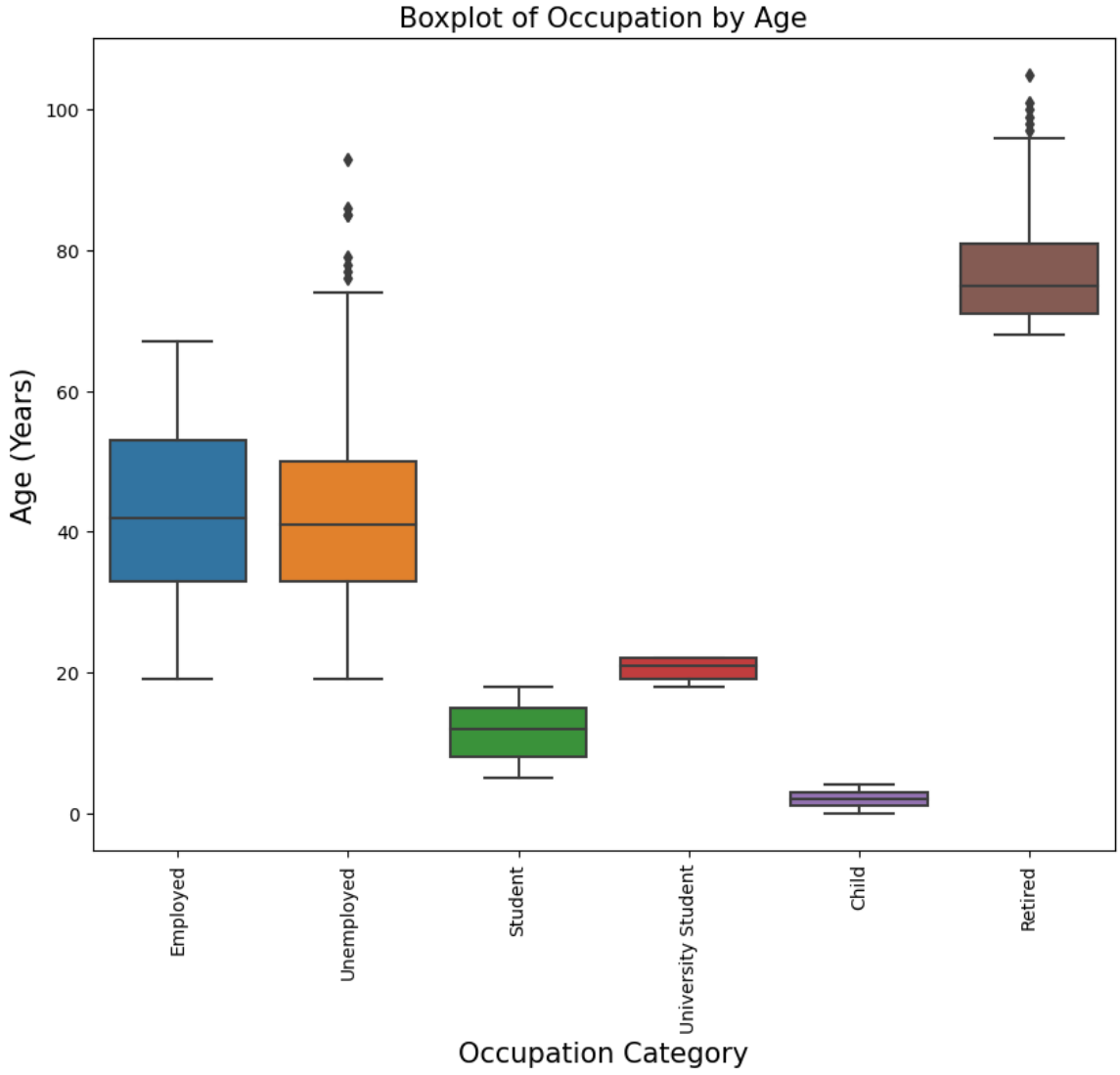


Fig 9. Distribution (Boxplot) of Occupation by Age

Based on the boxplot in **Fig 9** and summary statistics presented, it can be observed that the average age of retired individuals is 76 years. In general, most individuals are employed at an older age (43 years), and also unemployed at (42 years), this makes them get retired at an older age (76 years) as well.

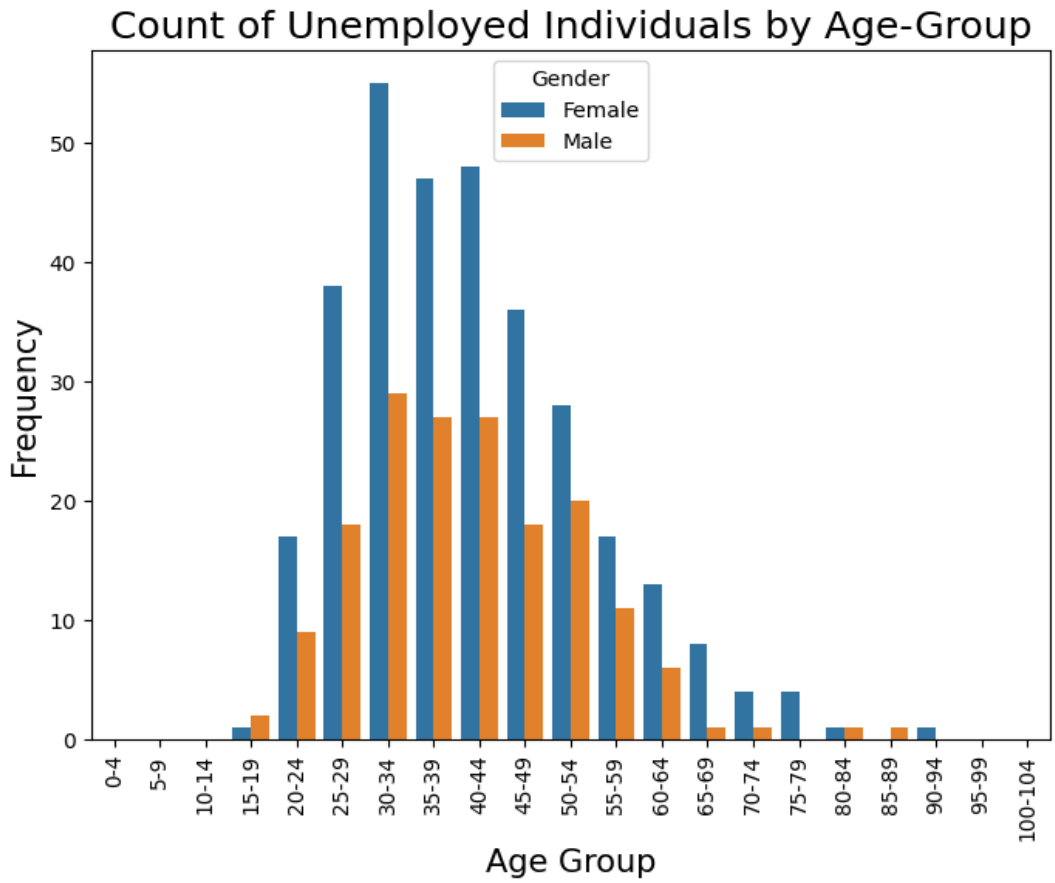


Fig 9. Counts of Unemployment by Age-Group

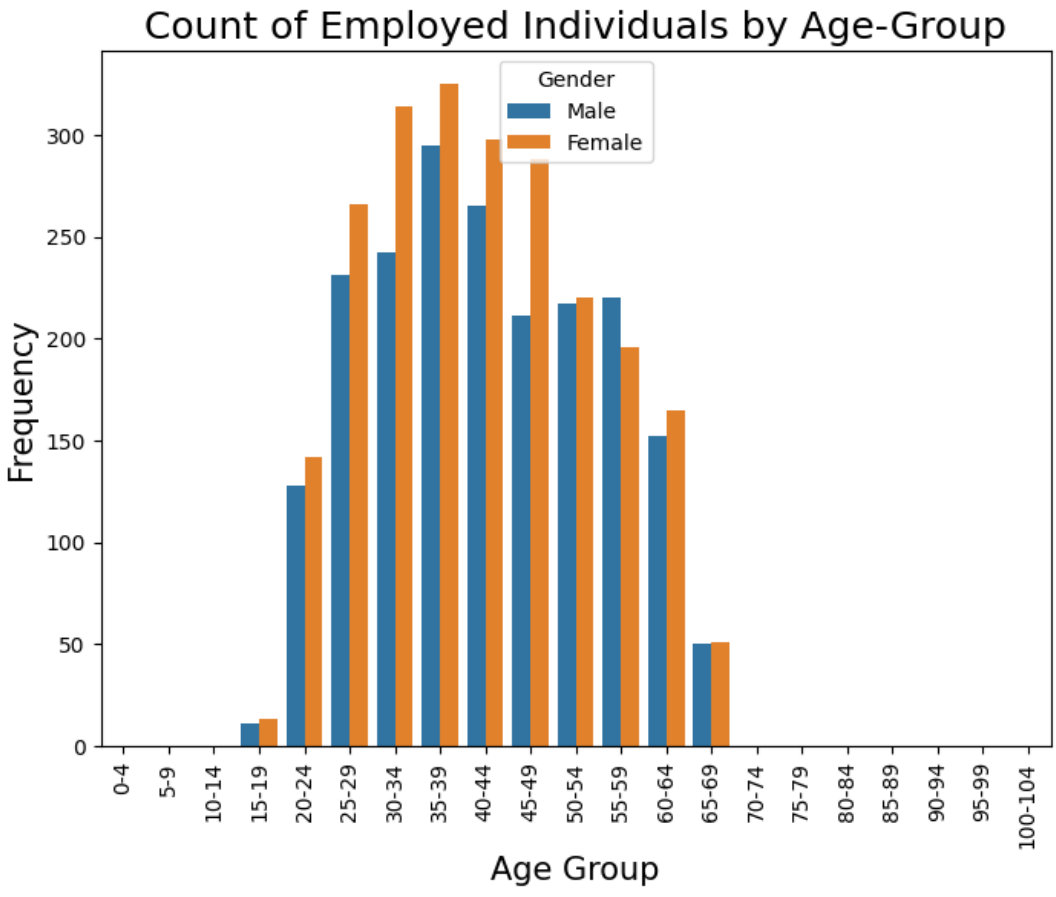


Fig 10. Counts of Employment by Age-Group

According to the chart provided in **Fig 9.**, the highest number of unemployed individuals belongs to the age bracket of 30 to 34, and the figures decrease with increasing age. Additionally, the number of unemployed females surpasses that of males in all age groups. However, the chart provided in **Fig 10.** Indicate that the age group of 35-39 has the highest number of employed individuals among both males and females. Additionally, females have a higher rate of employment than males in the age range of 15-54, and females are mostly employed than the males.

The employment rate is 89.79%, which indicate that a significant majority (89.79%) of people have jobs or are self-employed, while (10.21%) are currently unemployed and actively seeking employment.

Marriage and Divorce Rate

```
Higher STD for widowed due to age range and farthest away from the 25% percentile
```

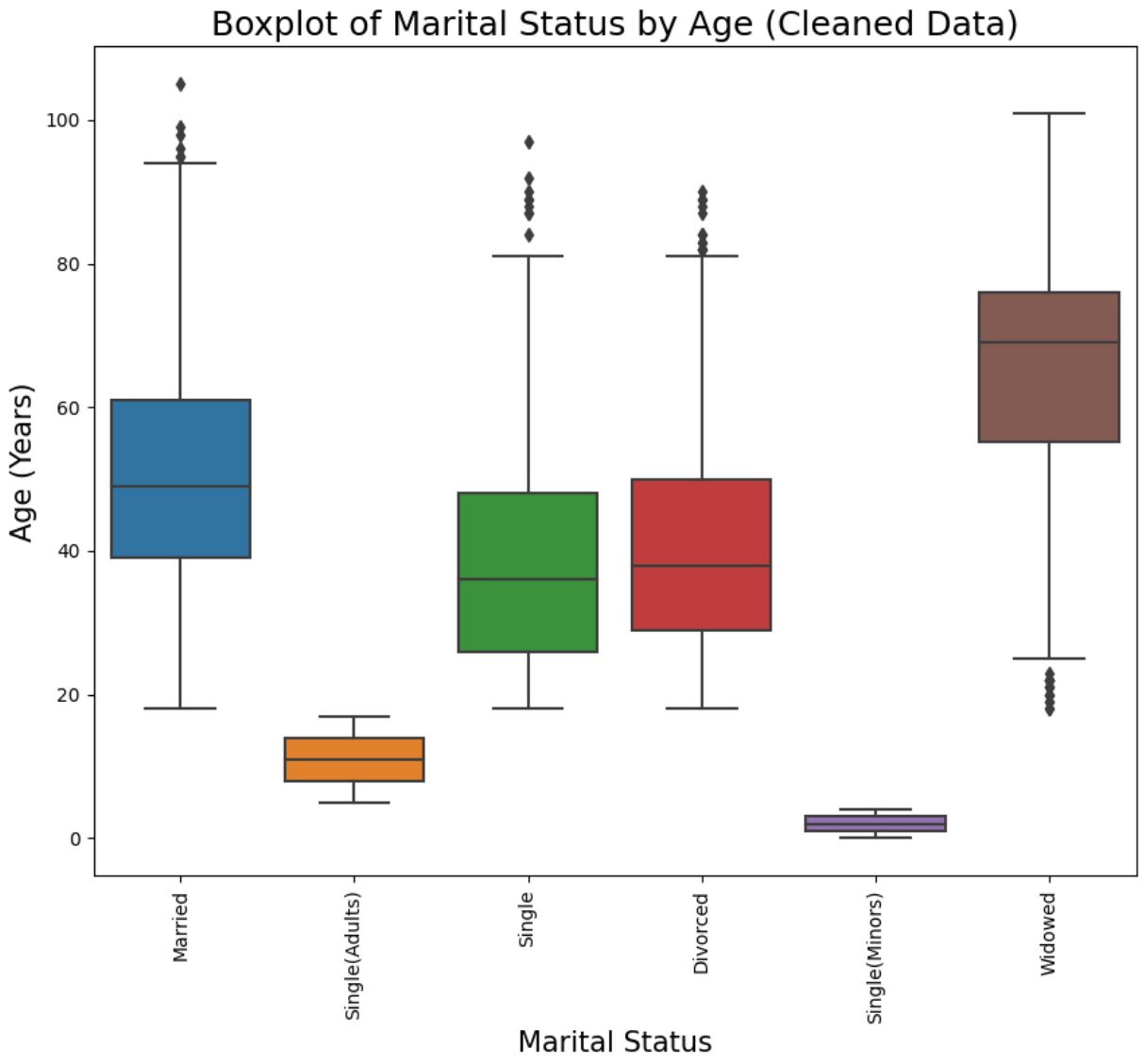


Fig 11. Boxplot/Distribution of Marital Status by Age

Marital_Status	count	mean	std	min	25%	50%	75%	max
Divorced	737.0	40.778833	15.879597	18.0	29.00	38.0	50.0	90.0
Married	2230.0	50.559193	16.101792	18.0	39.00	49.0	61.0	105.0
Single	2695.0	37.823006	13.776379	18.0	26.00	36.0	48.0	97.0
Single(Adults)	1517.0	11.059328	3.662426	5.0	8.00	11.0	14.0	17.0
Single(Minors)	447.0	2.100671	1.400248	0.0	1.00	2.0	3.0	4.0
Widowed	278.0	63.208633	19.429211	18.0	55.25	69.0	76.0	101.0

According to the chart provided in **Fig 11.,** and summary statistics presented**,** the widowed population is the oldest and least group among the married, divorced, and various single populations. Divorce rate per 1000 persons is approximately 93 (93.2) and marriage rate per 1000 persons is approximately 375 (375.38) of the total population.

```
Higher STD for widowed (males and females) due to age range and farthest away from
                                the 25% percentile
```

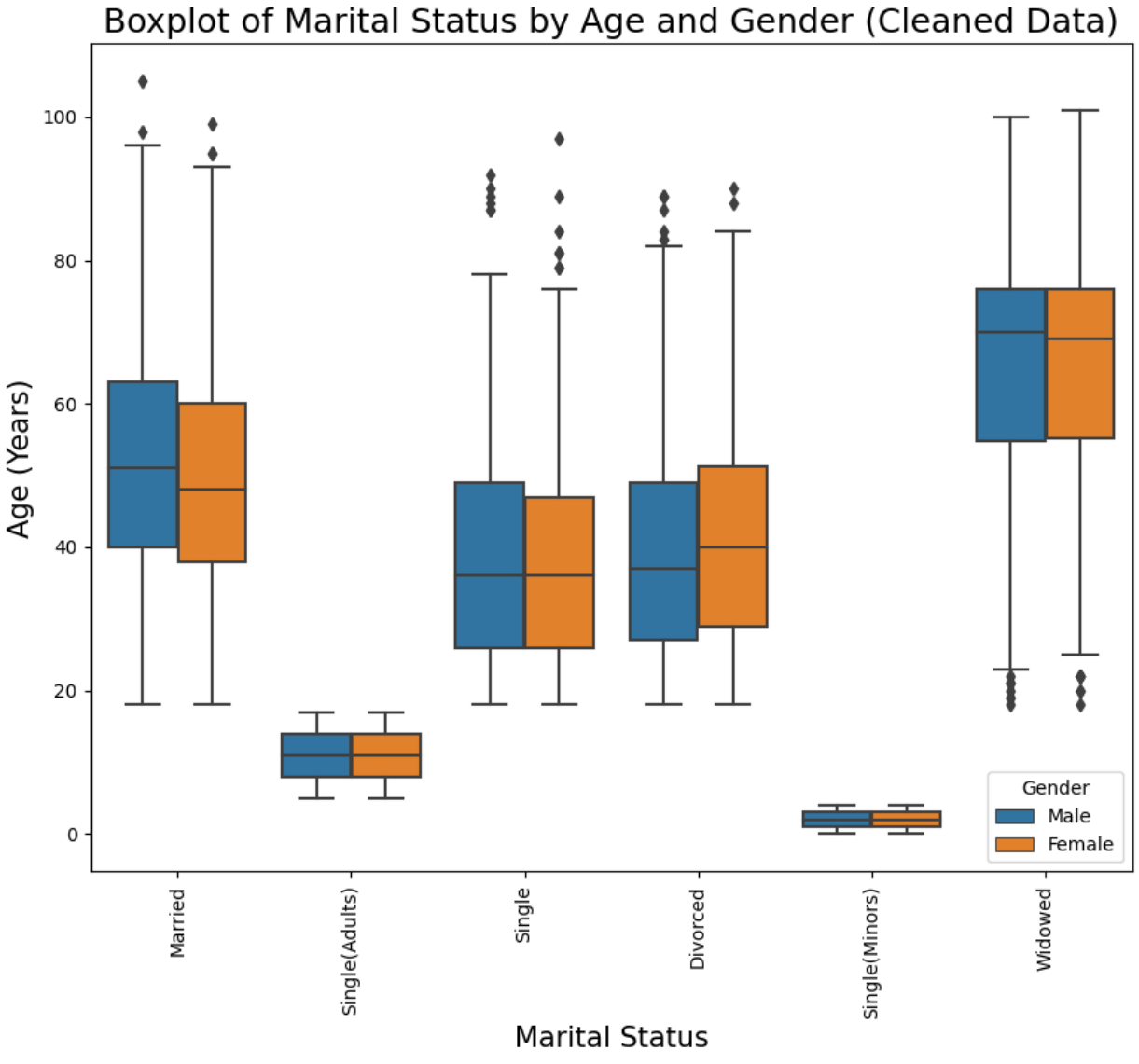


Fig 12. Boxplot/Distribution of Marital Status by Age and Gender

Marital_Status	Gender	count	mean	std	min	25%	50%	75%	max
Divorced	Female	452.0	41.460177	15.636551	18.0	29.00	40.0	51.25	90.0
	Male	285.0	39.698246	16.226614	18.0	27.00	37.0	49.00	89.0
Married	Female	1115.0	49.519283	16.035596	18.0	38.00	48.0	60.00	99.0
	Male	1115.0	51.599103	16.107856	18.0	40.00	51.0	63.00	105.0
Single	Female	1449.0	37.725328	13.399583	18.0	26.00	36.0	47.00	97.0
	Male	1246.0	37.936597	14.206565	18.0	26.00	36.0	49.00	92.0
Single(Adults)	Female	731.0	11.084815	3.756773	5.0	8.00	11.0	14.00	17.0
	Male	786.0	11.035623	3.574681	5.0	8.00	11.0	14.00	17.0
Single(Minors)	Female	224.0	2.133929	1.395030	0.0	1.00	2.0	3.00	4.0
	Male	223.0	2.067265	1.407814	0.0	1.00	2.0	3.00	4.0
Widowed	Female	170.0	63.235294	18.927546	18.0	55.25	69.0	76.00	101.0
	Male	108.0	63.166667	20.283274	18.0	54.75	70.0	76.00	100.0

Fig 12. Summary Statistics of Marital Status and Age Conditioned by Gender

According to the table in **Fig 12.** and summary statistics presented, the number of divorced females is higher than males i.e., the population of divorced females is approximately 61% (61.3%) and that of males is approximately 39% (38.7%), while the number of married females is equal to that of males, also women in the town divorce later than the men.

Birth and Death Rate

The **crude birth rate** is the number of live births per 1000 of the total population, mathematically expressed by.

$$\text{Crude Birth Rate} = \frac{Number\ of\ live\ births\ (Age\ 0\ years)}{Total\ Population}\ x\ 1000$$

The calculated crude birth rate of the population is 9.74, which indicate that out of every 1000 persons in the town approximately 10 (9.84) are live births. A crude birth rate of 18 or fewer is considered to be low (Safeopedia, 2017). The crude birth rate of the population is low.

The **crude death rate** is the number of deaths per 1000 of the total population, mathematically expressed by.

$$\text{Crude Death Rate} = \frac{Number\ of\ deaths}{Total\ Population}\ x\ 1000$$

The average crude death rate of the population was obtained by assumption, whilst considering the aged population (individuals in the town most likely to pass away) in age groups (56-60, 61-65, till the oldest age groups) per annum. The average crude birth rate of the population is 3.8, this signifies that, on average, 3.8 people out of every 1,000 people in the population die in a given time period, usually a year. In general, an average rate of 3.8 suggests that the population is relatively healthy and has a moderate to low mortality rate.

RECOMMENDATIONS

Due to the low occupancy rate and large number of commuters in the town, it is pertinent to construct a train station and consider investing in low-density housing. Low-density housing could prove advantageous for families, divorced individuals with young children who are looking to downsize, as well as lodgers moving into the town. Additionally, the train station would be beneficial for people who need to travel to adjacent cities for work, school, or religious activities.

Most people in the town are christians, it is necessary for the government to build a church in the town. This would also alleviate the commuters rate in the town.

Conclusively, investing in old age care should be a priority since very few individuals suffer from infirmity and the population is aging, and more people may suffer from health issues in the future. Prior to the ageing population growing and necessitating more care, it is imperative to invest in these services.

REFERENCES

Gov.uk (No Date) *Marriage and Civil Partnership (Minimum Age) Act (2022)*
Available online: https://www.legislation.gov.uk/ukpga/2022/28/enacted [Accessed 26/04/2023]

Marriage Act (1949) Section 3
Available online: https://gro-extranet.homeoffice.gov.uk/wp-content/uploads/2013/07/Marriage-Act-1949-loadedJune2015.pdf#:~:text=%283%29%20Any%20such%20marriage%20as%20is%20mentioned%20in,attained%20the%20age%20of%20twenty-one%20at%20the%20time [Accessed 26/04/2023]

National Society for the Prevention of Cruelty to Children (2020) *Moving Out*
Available online: http://www.nspcc.org.uk/keeping-children-safe/in-the-home/moving-out/ [Accessed 26/04/2023]

RoboCade (2023) *Anti-Sith bill*
Available online: https://starwars.fandom.com/wiki/Anti-Sith_bill?oldid=12288401 [Accessed 26/04/2023]

Safeopedia (2017) *Crude Birth Rate*
Available online: https://www.safeopedia.com/definition/2506/crude-birth-rate#:~:text=A%20crude%20birth%20rate%20of,babies%20were%20born%20last%20year [Accessed 26/04/2023]

United Nations Convention on the Rights of the Child (1990) Article 14
Available online: https://www.unicef.org.uk/wp-content/uploads/2010/05/UNCRC_united_nations_convention_on_the_rights_of_the_child.pdf [Accessed 26/04/2023]