NAME: OBARAKPOR TEGA

STUDENT ID: 202120109

CENSUS PROJECT REPORT

INTRODUCTION

This report analyses a census dataset for an imaginary small town sandwiched between two larger cities in order to make suggestions to the government on future growth and investments. The dataset comprises demographic and socioeconomic statistics for the town's population, such as age, gender, address, occupation, religious affiliations, marital status with the intention of gathering information such as population density, employment trends, religious affiliations and trends, divorce and marriage rates, occupancy levels of housing, death and birth rates, and likely commuters. The first section of this report goes through the data cleaning and the techniques utilised to assure the dataset's dependability and accuracy. The next parts analyze the data to evaluate the best use of an unoccupied parcel of land and to choose which area of investment the community should prioritise.

DATA CLEANING

Before any meaningful analysis could be carried out, errors in the data collected needed to be addressed through a process of data cleaning. The dataset was in CSV format and a Jupyter notebook was used for this process. Steps taken included checking for duplicates, ensuring appropriate data types for respective columns, finding and deciding how to deal with Nan values, inspecting and cleaning columns individually.

The dataset was scanned for duplicates and one entry was removed. The Age column had some ages recorded in decimal places, so all float values were converted to int values to ensure correct age in years. Inspection of the unique values in the Gender column showed one blank gender value, which was located and resolved through a deduction process based on the person's marital status.

The Religion column had a large number of Nan values and some strange entries like Jedi, Sith (a typo for Sikh), and Nope (likely indicating None). All Nan values were changed to Not Available because religion is faith based and independent, and there is no logical analysis to assign religion as people within one household sometimes have different religions, while the Sith values was changed to Sikh, then Nope, Jedi, and Undecided values were changed to Not Available.

The Marital Status column had 2484 Nan values, all of which were discovered to belong to people under 18 years which is the legal marrying age in the UK (gov.uk, 2023). These were changed to Ineligible instead of single to distinguish minors from single adults. Upon further investigation, it was discovered that one of the marriages in the dataset involved a wife who was below the legal age for marriage while the husband was legal. Although dropping both records from the dataset was considered, it was found that the husband is a university student and his inclusion was important for future analyses on commuting or building a

university which involved the limited number of university students. Therefore, the marriage was voided and both individuals were categorized as single and ineligible. This decision was made based on the fact that it would have a lesser impact on the marriage rate than dropping them would have on the commuting and university analyses. A blank value was also changed to Married based on reviewing neighboring rows.

The Infirmity column had 10 blank values and 10 values for disabled without any specifics, both of which were changed to unknown infirmity and unknown disability, respectively. The occupation column had no inconsistencies, while two blank values were found in the Relationship to Head of House column and accurately assigned Daughter and Husband.

Finally, three blank values were found under the Surname column and were located along with neighboring rows and assigned the correct surnames. Overall, these cleaning steps ensured the data was ready for analysis.

DATA ANALYSIS

POPULATION DEMOGRAPHICS

To aid the analysis of the population demographics, a 5-year age band has been added to the data for the population pyramid.
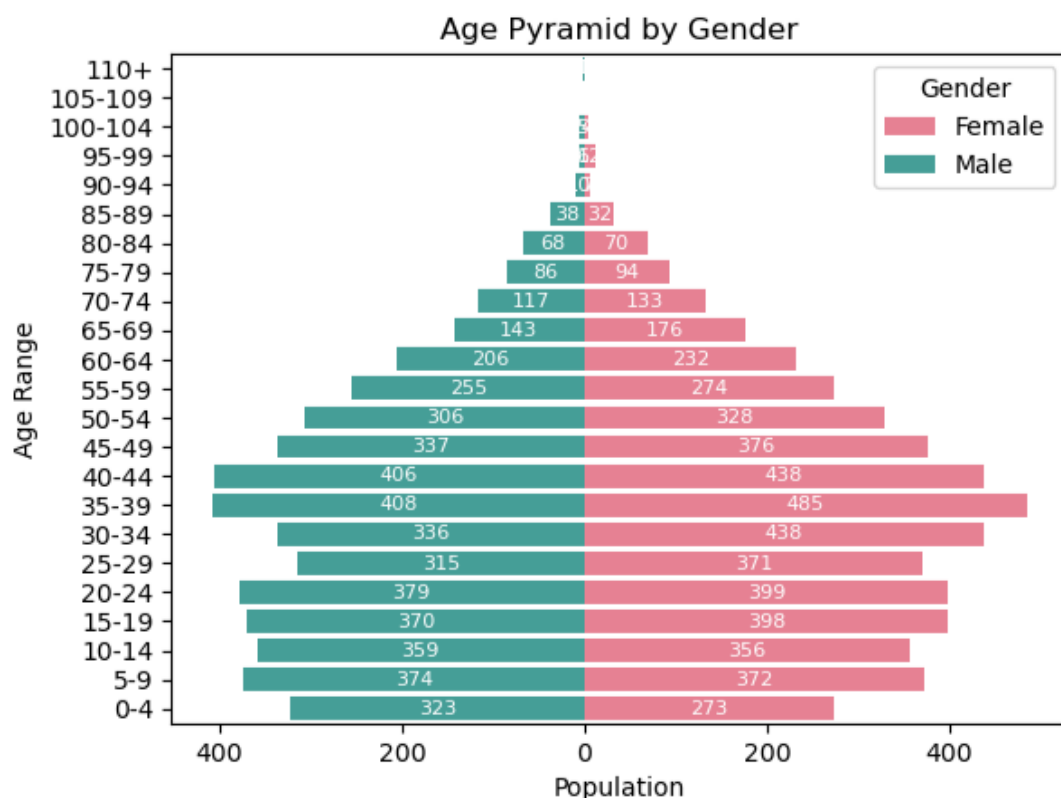

Figure 1: Population Pyramid from Jupyter Notebook

The population of the town is split evenly between genders, with slightly more females than males. Largest group being 35-44, and 85 upwards being the smallest.

There are more male than female children below 15, but there are more females than males in age groups above 15, indicating that there are more middle-aged females than males in the town. This may suggest that resource allocations or policies for middle-aged people should be slightly female-inclined.

The narrow top of the pyramid suggests that older people tend to migrate out of the town after retirement as deducing death is inconsistent with the life expectancy in the UK currently set at 79.3 years for men and 83.1 years for women (ONS, 2021).

If migration becomes minimal, the data suggests that there will be an increasing number of retired people in the future, as the population count for age groups close to 65 is relatively high. Also, there is a significant decrease in the birth rate in the last 5 years, indicating that there will be fewer school-aged children in the future.

A limitation in analyzing the population demographics of the town is the lack of historical data for comparison, as well as limited information on actual death rates and live births, and migration patterns. This can make it challenging to obtain accurate insights into the population trends and changes over time.

UNEMPLOYMENT TRENDS
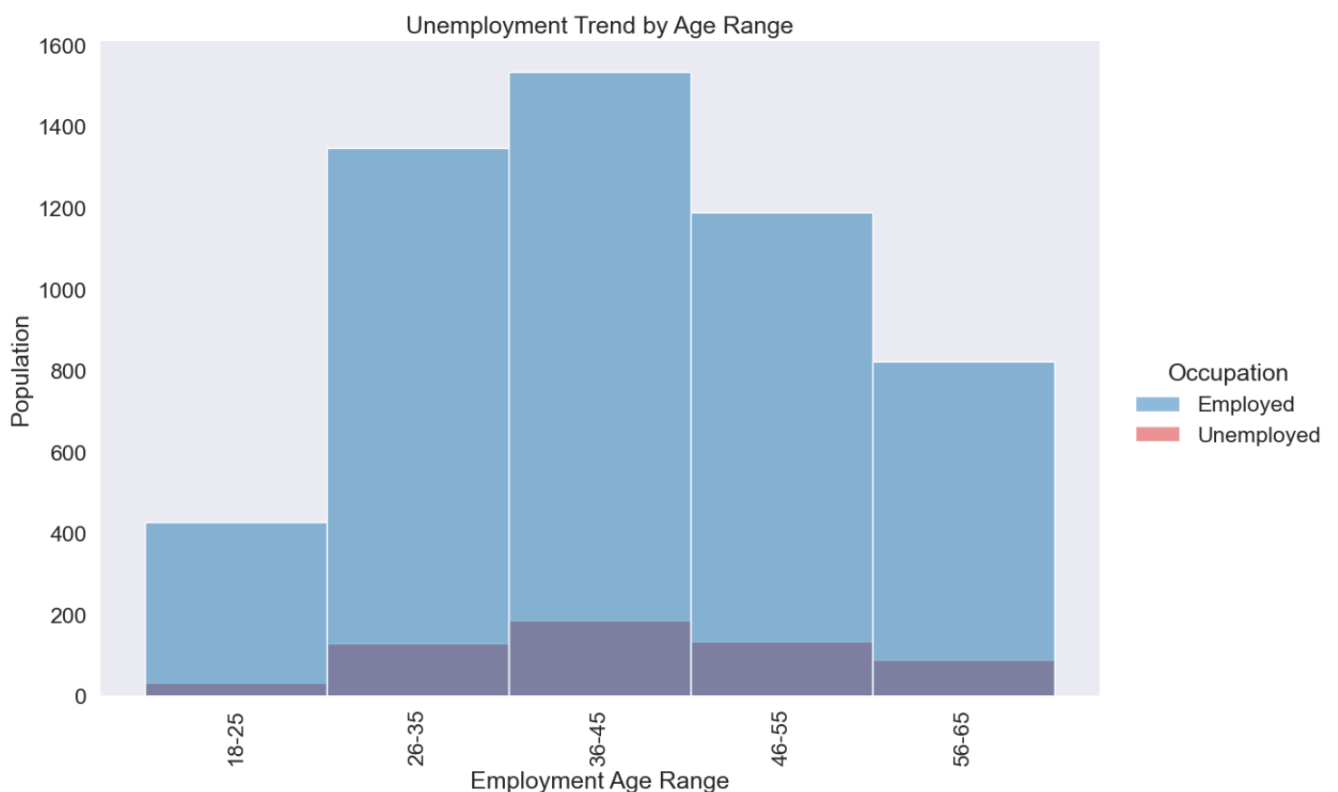
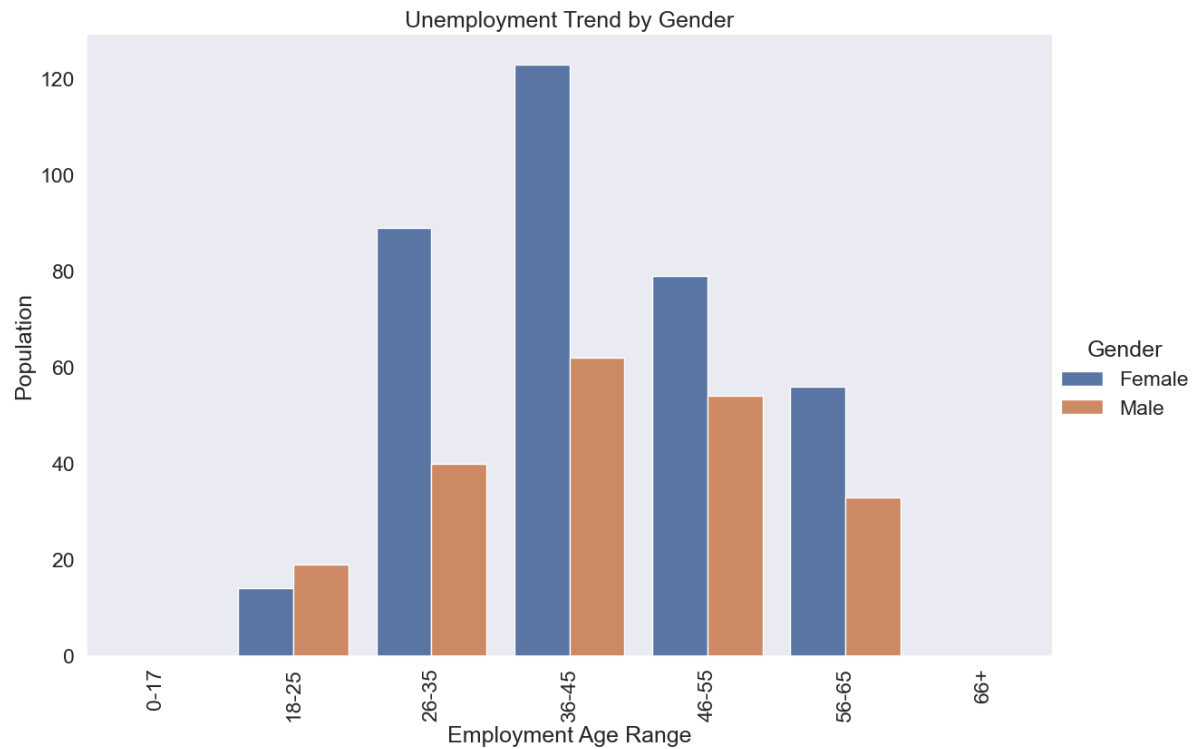*Figures 2: Unemployment by Age Group from Jupyter Notebook*

Figure 3: Unemployment Trend by Gender from Jupyter Notebook

UNEMPLOYMENT BY AGE GROUP

Occupation	Employed	Unemployed
Employment Age Range		
18-25	92.81	7.19
26-35	91.25	8.75
36-45	89.24	10.76
46-55	89.93	10.07
56-65	90.22	9.78

UNEMPLOYMENT BY GENDER

Gender	Female	Male
Employment Age Range		
18-25	42.42	57.58
26-35	68.99	31.01
36-45	66.49	33.51
46-55	59.40	40.60
56-65	62.92	37.08

The analysis of the data reveals that middle-age groups have the highest unemployment rate compared to the youngest and oldest employment age groups, it also shows higher unemployment for females. The largest difference in unemployment rates by gender is in the 26-35 and 36-45 groups respectively. These findings underscore the need for female-oriented programs, resources, and policies in those age groups to bridge the gender unemployment gap.
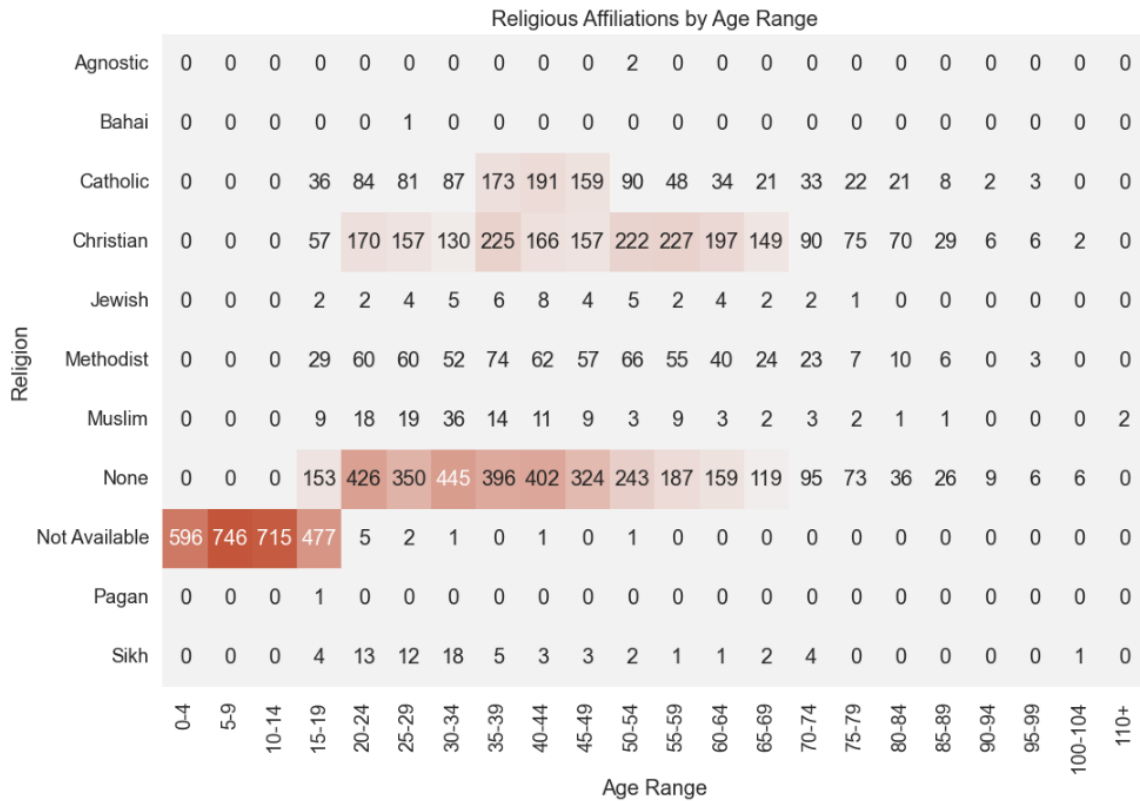
RELIGIOUS TRENDS AND AFFILIATIONS

Religious Affiliations by Age Range

Religion	0-4	5-9	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80-84	85-89	90-94	95-99	100-104	110+
Agnostic	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0
Bahai	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Catholic	0	0	0	36	84	81	87	173	191	159	90	48	34	21	33	22	21	8	2	3	0	0
Christian	0	0	0	57	170	157	130	225	166	157	222	227	197	149	90	75	70	29	6	6	2	0
Jewish	0	0	0	2	2	4	5	6	8	4	5	2	4	2	2	1	0	0	0	0	0	0
Methodist	0	0	0	29	60	60	52	74	62	57	66	55	40	24	23	7	10	6	0	3	0	0
Muslim	0	0	0	9	18	19	36	14	11	9	3	9	3	2	3	2	1	1	0	0	0	2
None	0	0	0	153	426	350	445	396	402	324	243	187	159	119	95	73	36	26	9	6	6	0
Not Available	596	746	715	477	5	2	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0
Pagan	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Sikh	0	0	0	4	13	12	18	5	3	3	2	1	1	2	4	0	0	0	0	0	1	0

Figure 5: Heatmap showing religious affiliations across different Age groups

	Frequency	Percentage
None	3455	34.150440
Not Available	2544	25.145794
Christian	2135	21.103094
Catholic	1093	10.803598
Methodist	628	6.207374
Muslim	142	1.403578
Sikh	69	0.682020
Jewish	47	0.464565
Agnostic	2	0.019769
Pagan	1	0.009884
Bahai	1	0.009884

Figure 6: Frequency distribution of Religions

The town is fairly irreligious, especially among the younger generation, with approximately one-third of the population falling under the "None" category. The "Not Available" category, which mainly includes children under 18, is the second largest group, making it unclear what religious affiliations these individuals may have in the future. The "Christian" and "Catholic" categories are the second and third most common religions respectively, but their frequencies are relatively low compared to the "None" category, indicating that these religions may be declining in popularity, especially the Christians with the bulk of their population being above 45. The "Muslim" and "Sikh" while few, may be growing as their

population is young. The "Agnostic", "Pagan", and "Bahai" categories have very low frequencies, suggesting that these religions have a small number of adherents.
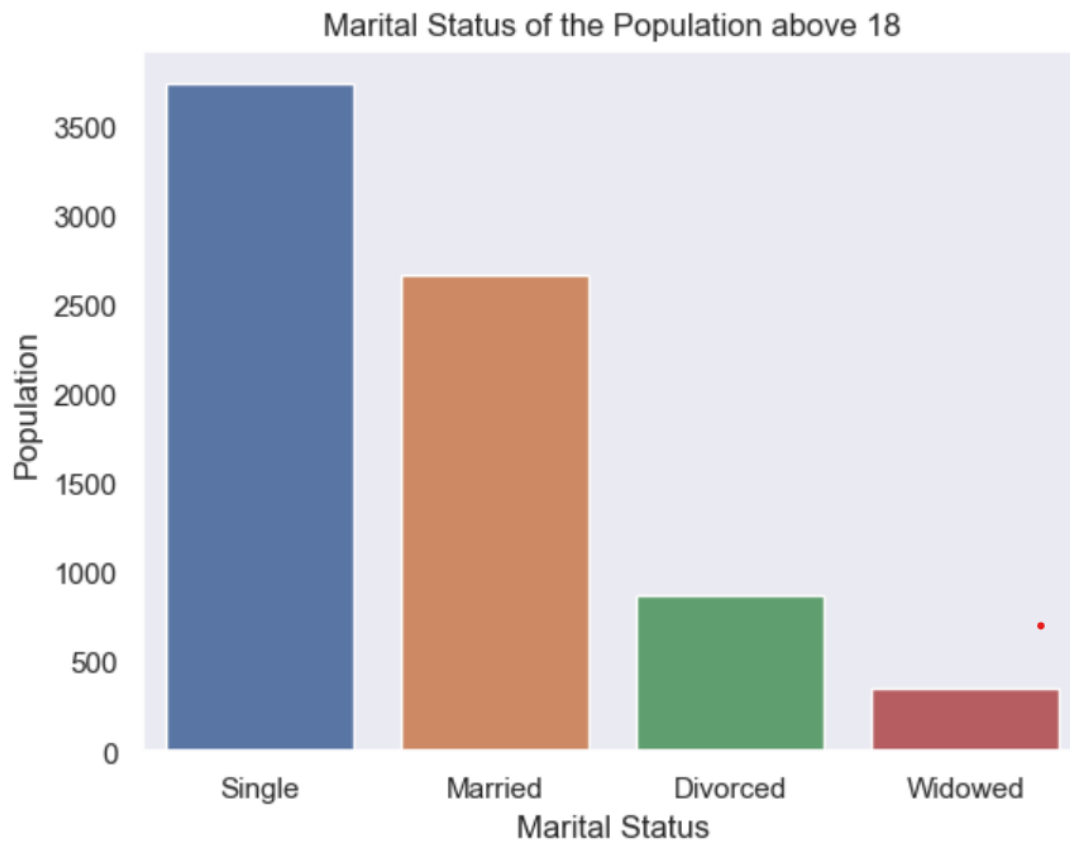
DIVORCE AND MARRIAGE RATES

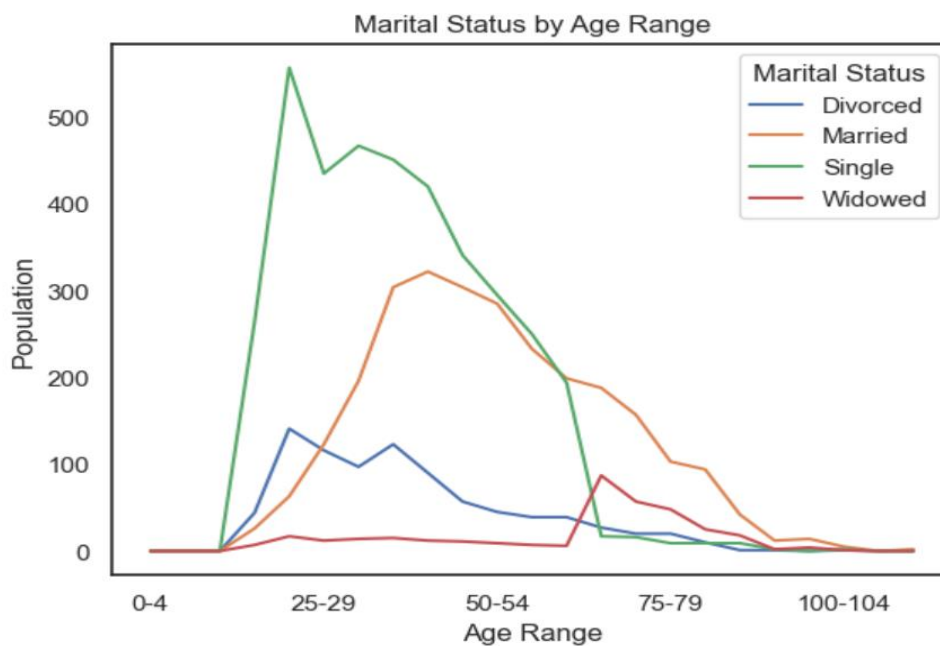*Figure 4: Divorce and Marriage rates*

Figure 5: Marital Status by Age Group

```
           Count  Frequency
Single      3738  48.971571
Married     2672  35.005895
Divorced     871  11.410979
Widowed      352   4.611555
```

In the town, the largest portion of the population is made up of single individuals, accounting for almost 49% of the total population, which suggests that the town may have a significant number of young adults or individuals who are not currently in long-term relationships. The majority of the single population falls within the early twenties age group, but this category declines as individuals get older, indicating that people tend to find companionship as they age.

The second largest group in the town is married individuals, comprising around 35% of the population. The bulk of the married population falls within the age range of 35 to 50, with a slight decline in numbers as individuals get older. This may suggest that individuals in this town tend to marry during their middle years.

The proportion of divorced individuals in the town is also notable, accounting for 11% of the population. The highest rate of divorces occurs among individuals in their early twenties, followed by the late thirties, and then a decrease in likelihood of divorce as individuals get older. This suggests that relationship dissolution is not uncommon, particularly among young people, and policymakers may benefit from understanding the underlying reasons for divorce and providing resources and support for those going through this difficult transition.

The proportion of widowed individuals in the town is relatively low, accounting for only 4.6% of the population. There are very few widows between the ages of 18-50, with a sharp increase in numbers around 65-69 years, followed by a gradual decline in older age groups. This along with the low numbers of infirmity recorded suggests that the town has a healthy population.

Overall, the data indicates that marriage is not a top priority for many residents or that there are barriers to forming long-term partnerships in the town. This could also be used as an indicator for housing policies geared towards providing high density housing as there would naturally be a demand for smaller homes
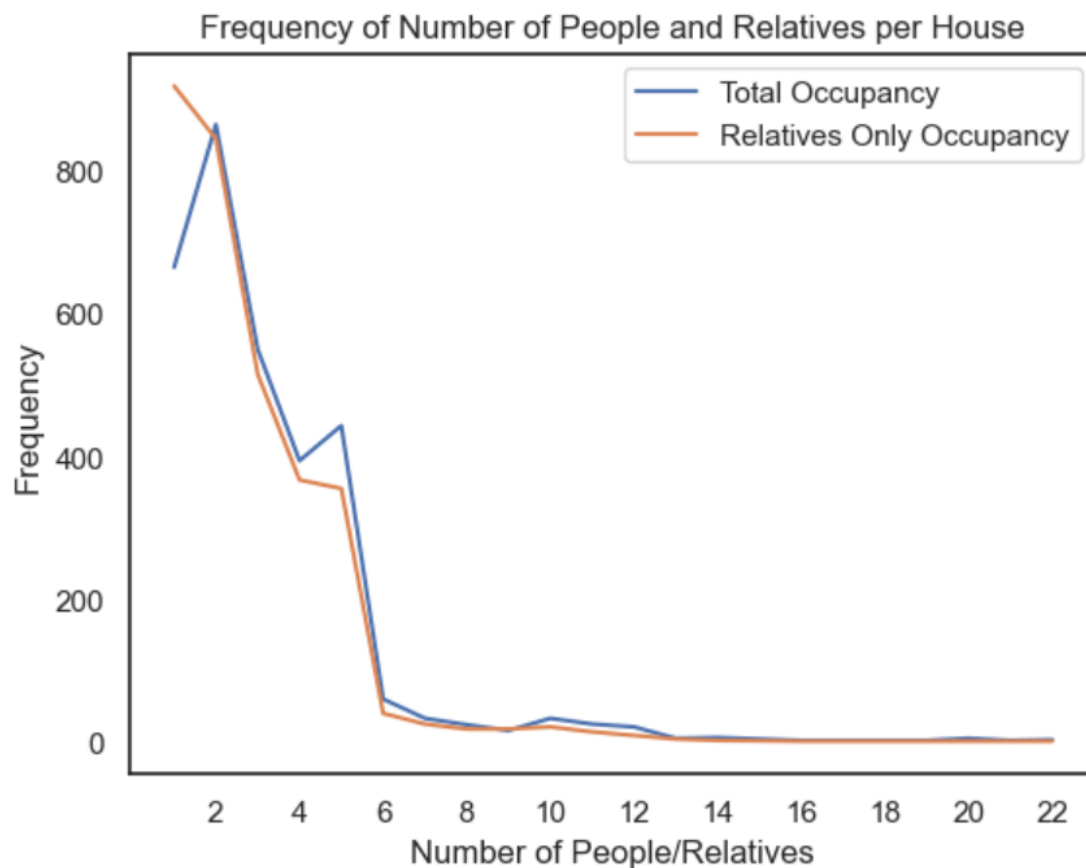
HOUSING OCCUPANCY LEVELS

Figure 6: Plot showing the number of occupants per house against the number of just relatives per house

The analysis showed that the distribution of the number of people per house is quite varied. There are houses with as few as one resident and as many as 20 residents. However, the most common number of people per house is two and one. And as seen in the line plot above, this changes to one per house when unrelated residents are removed.
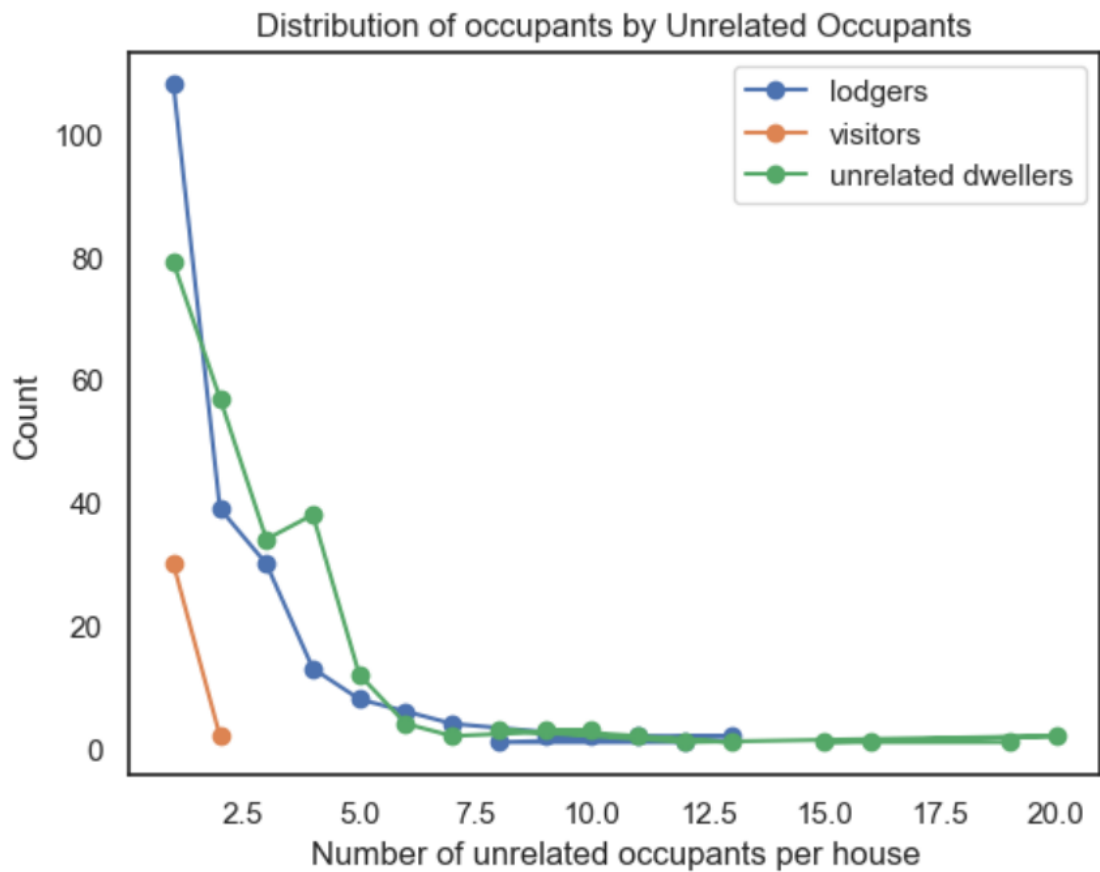
Figure 7: Distribution of Unrelated occupants per house

The distribution of the number of lodgers, unrelated occupants and visitors per house indicates that they are relatively few in the population, with most houses having only one or two. This suggests that the housing needs of the population are primarily driven by individual needs followed by family needs. However, it is interesting to note that there are a few houses with a significant number of residents who are not related to the head of the house.
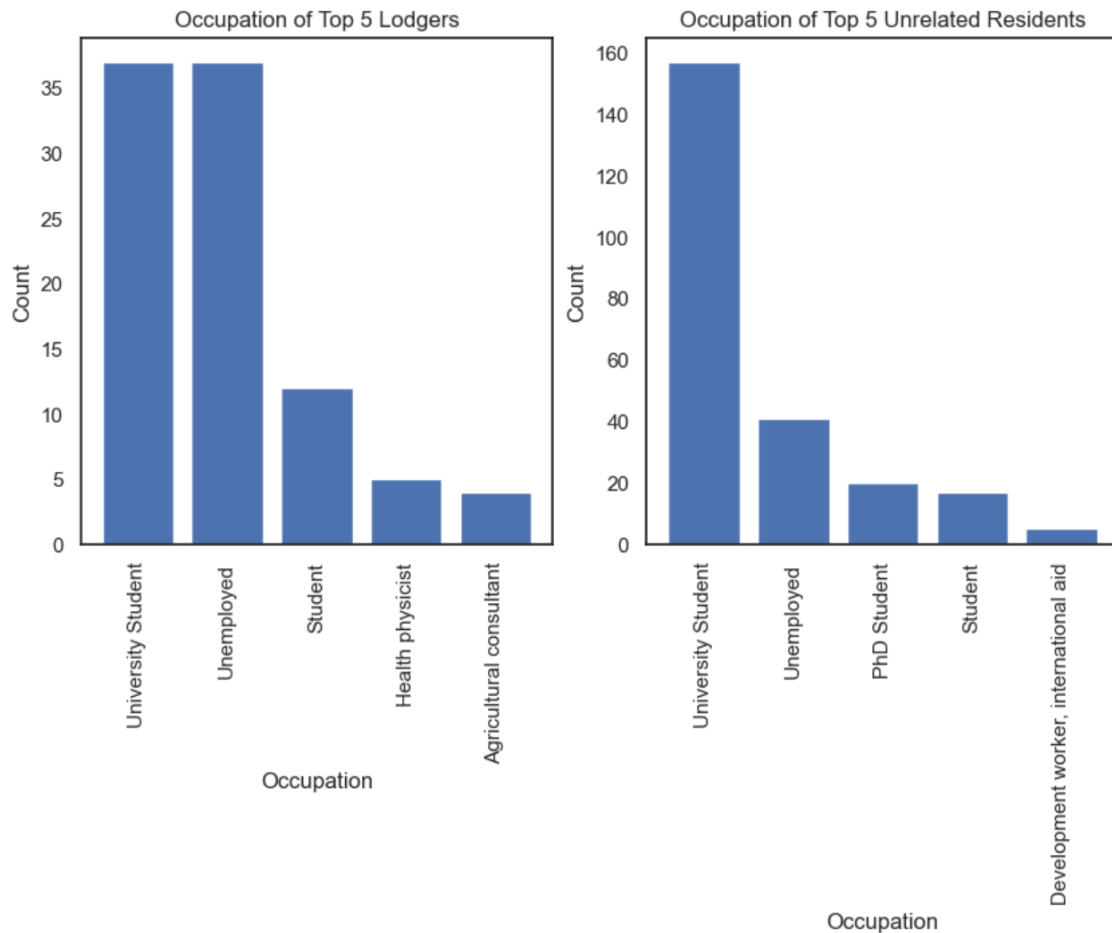
Figure 8: Occupation of lodgers and unrelated residents

The fact that a high number of these unrelated residents are university students also indicates that there is a need for student housing or accommodations that cater to younger individuals who are not necessarily part of a family unit.

Summarily, the prevalence of one- or two-persons households suggests that there may be a need for smaller homes or apartments that cater to individuals or small families and this is even more evident when the analysis gotten from the marital status of the population is taken into account as well.

One limitation is that while the data suggests that there is a range of occupancy levels in the population, it is difficult to make a definitive determination of whether existing housing is being under or over-used without more information about local norms and standards.

UNIVERSITY STUDENTS AND COMMUTERS

There are 669 university students living in the town and it as assumed that all of them are commuting out of town unless there is the possibility for online study. As this is a modest town, it is also assumed there are no airports or military base, therefore doctors, surgeons, lecturers, professors, researchers, military and airline workers are also classified among

those commuting out of town for work and they would make up the total range of the population commuting out of town.

The total number of likely commuters in the town is estimated at 1071 which makes up 10.58% of the total population

BIRTH AND DEATH RATE ANALYSIS

$$Birth\ Rate = \frac{\text{Number of Live Births}}{\text{Total Population}}\ X\ 1000$$

The birth rate is calculated by assuming all children who haven't reached their first birthday as the number of live births since they were born within the last year, this number is then divided by the total population then multiplied by 1000 to get the birth rate per thousand.

$$Death\ Rate = \frac{\text{Decline}}{\text{Total Population}}\ X\ 1000$$

Decline = ∑ difference in age band over 65

The death rate is calculated by estimating deaths as the summation of the difference in age bands for those over 65. The summation is divided by the total population and multiplied by 1000 to get the estimated death per thousand.

The current birth rate for the town is estimated at 10 births per 1000, while the previous birth rate is estimated at 13 births per 1000, which indicates a fall of 3 children per 1000 over 5 years. The death rate for the town is estimated at 6.31 deaths per 1000. It is acknowledged that these results are only an estimated calculation as the data set provides no information on migration as well, which could also be a determining factor of the values assessed.

CONCLUSION

(Wahab & Falola, 2019) argue that planning at the local government level is very important for development and it should be flexible and involve everyone who will be affected by the development. From the analysis of the data, building high-density housing would be the most impactful and inclusive decision as it is based on the prevalence of single adults, one or two-person households, young age divorce, university students, lodgers, and an expected increase in young adults. Building high-density housing will provide additional housing options for the growing population while utilizing less land. For other options considered, establishing a train station could be beneficial for commuters, but they only make up 10% of the population and insufficient information regarding commuter patterns and demand prevents accurate evaluation. Another religious building may not be necessary, as there is already a Catholic place of worship in the town and the population of growing religions is too small. Building emergency medical facilities is an option that includes the entire population, but the land may not be optimally used as the center would only treat minor injuries and not function as a full hospital.

(Lazaro & Maria 2011) stressed the importance of investing in social programs and economic development in urban development and sustainability. It is recommended that the government invest in employment and training, particularly for young female adults who are most affected by unemployment. Providing training for new skills can decrease the unemployment rate and ensure that the population is equipped to participate in the growing economy. General infrastructure development, such as waste collection and road maintenance, is also essential to support the growing population's needs. Although funding for school and elderly care is important, it is not considered a priority, given the expected decrease in school-aged children and the significant increase in the retired population, which is still about a decade away.

REFERENCES

Gov.uk (2023) *Legal age of marriage in England and Wales rises to 18*. Available Online: https://www.gov.uk/government/news/legal-age-of-marriage-in-england-and-wales-rises-to-18 [Accessed 26/04/2023]

Lazaro, T. & Maria, J. (2011). *Urban Development and Sustainability*. European Journal of Geography, 2(1), 38-50. Retrieved from https://www.researchgate.net/publication/283462728_URBAN_DEVELOPMENT_AND_SUSTAINABILITY

Office for National Statistics (2021) *National life tables – life expectancy in the UK: 2018 to 2020*. Available Online: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/lifeexpectancies/bulletins/nationallifetablesunitedkingdom/2018to2020 [Accessed 26/04/2023]

Wahab, B. & Falola, O. (2019). *Local Government Administration and Physical (Development) Planning in Africa*. In O. Falola (Ed.), Contemporary Issues in Local Government Administration in Africa (pp. 39-58). Lambert Academic Publishing. Retrieved from https://www.researchgate.net/publication/343306266_Local_Government_Administration_and_Physical_Development_Planning_in_Africa