CENSUS PROJECT REPORT

INTRODUCTION

The census process involves gathering, organizing, analyzing, assessing, publishing, and disseminating statistical information on the population, housing, and geographical location. It provides a comprehensive snapshot of society at a specific time, including ongoing changes that cannot be captured by other data collection methods. The census is a valuable source of information for comparing and projecting demographic, social, and economic trends.

The dataset reveals the existence of a town of moderate size, situated between two larger cities linked by highways. Although there is no university in the town, students live there and commute to nearby cities for their education. The goal is to provide recommendations on suitable investments and what to construct on an empty plot of land. To achieve this objective, the report will employ various techniques, including data cleaning, exploratory data analysis, and data visualization.

DATA CLEANING

The biggest challenge with data is that data are always filled with errors because data in its raw form is always poorly managed. The process of data cleaning is necessary to ensure a consistent dataset that provides accurate information. You can find the detailed process of cleaning the dataset in the attached Jupiter notebook. The dataset contains 6000 Data and 11 Features. Each feature in the dataset was reviewed for errors, missing data, and inconsistent information.

A blank space was discovered in the first name column and it was replaced with ‘Unknown’.

- Age contained a missing record and it was filled with the median Age. The Age column data type was an object and it was casted to its original data type which is an integer.
- Two households were removed from the dataset because the head of the household was under 18 years old, and they were divorced, which is not legally possible unless the marriage has lasted for at least a year (Gov.uk).One of the household relationships to the head of the house was modified.
- All the missing values for marital status were converted to "NA" because they were within the age range of 0 to 17, which comprises minors (NSPCC, 2022).

- There was only one missing record for occupation, and it was replaced by selecting the mode of occupation for individuals within the same age group. Occupations with ambiguous information, such as copy, make, sub and land, were classified as "Not Given." Additionally, individuals aged 65 above, as well as those who were unemployed, were reclassified as "Retired" since they have limited chances of being employable, even though the mandatory retirement age of 65 no longer exists in the United Kingdom (Gov.uk).
- The blank spaces in the infirmity column was replaced with ‘None’. The missing values and blank spaces in religion were replaced with ‘Unknown’.

FEATURE ENGINEERING

After completing the data cleaning stage, the dataset has been improved to provide more accurate and useful information. To enable further analysis and modeling, additional features have been added to the census data.

These features include:

- **Age grouping:** The age is separated into 10-year intervals to form an age pyramid.
- **Occupation categories:** The occupation is classified into six different categories, which are Child, Student, Employed, Unemployed, Retired, and unknown (for occupations that lack detailed information).
- **Skillset:** Using the Standard Occupation Classification (SOC, 2020) as a reference, the occupation is classified into three categories based on skill levels: highly skilled, skilled, and skilled-trade.
- **Commute:** The occupation is classified into two categories, commuters and non-commuters, based on their preference for particular skill sets.

DESCRIPTIVE ANALYSIS AND VISUALIZATION

Age

```
count    5995.000000
mean       36.818182
std        21.716820
min         0.000000
25%        19.000000
50%        35.000000
75%        52.000000
max       105.000000
Name: Age, dtype: float64
```

Value	Count	Frequency (%)
30-39	961	16.0%
20-29	924	15.4%
40-49	891	14.9%
10-19	827	13.8%
50-59	734	12.2%
0-9	674	11.2%
60-69	501	8.4%
70-79	281	4.7%
80-89	166	2.8%
90-99	25	0.4%

After cleaning the data, there are 5995 records of individuals with varying age groups, skills, religions, and other factors. The mean and the median age are closely the same which means the age of the population are almost symmetrically distributed. The oldest age in the population is 105 years. 16% of the population falls between 30 -39 age group, 15.4% of the population falls between the age of group of 20-29 and 14.9% falls between the age group of 40-49 which means we have a lot of working class and university students in the population.

Gender

Value	Count	Frequency (%)
Female	3145	52.5%
Male	2850	47.5%

The female population is more than the male population. 52.5% of the populations are female while 47.5% are male.

Marital Status

Value	Count	Frequency (%)
Single	2161	36.0%
Married	1684	28.1%
NA	1279	21.3%
Divorced	581	9.7%
Widowed	290	4.8%

The population has more single people than married people. 9.7% of the population are divorced, 4.8% are widowed and 21.3% are the population are minors (less than 18) which was replace with 'NA' during the data cleaning.

Occupation Category

Value	Count	Frequency (%)
Employed	3261	54.4%
Student	1471	24.5%
Retired	593	9.9%
Unemployed	337	5.6%
Child	319	5.3%
Unknown	14	0.2%

54.4% of the population are employed which account for more than half of the town's population followed by students (24.5%).

Religion

Value	Count	Frequency (%)
None	2105	35.1%
Christian	1406	23.5%
Unknown	1319	22.0%
Catholic	620	10.3%
Methodist	407	6.8%
Muslim	87	1.5%
Sikh	28	0.5%
Jewish	18	0.3%
Private	2	< 0.1%
Orthodoxy	1	< 0.1%
Other values (2)	2	< 0.1%

Christian

Most of the populace is not affiliated with any particular religion. Christianity is the only dominating religion in the town which means 23.5% of the entire populations are Christians.

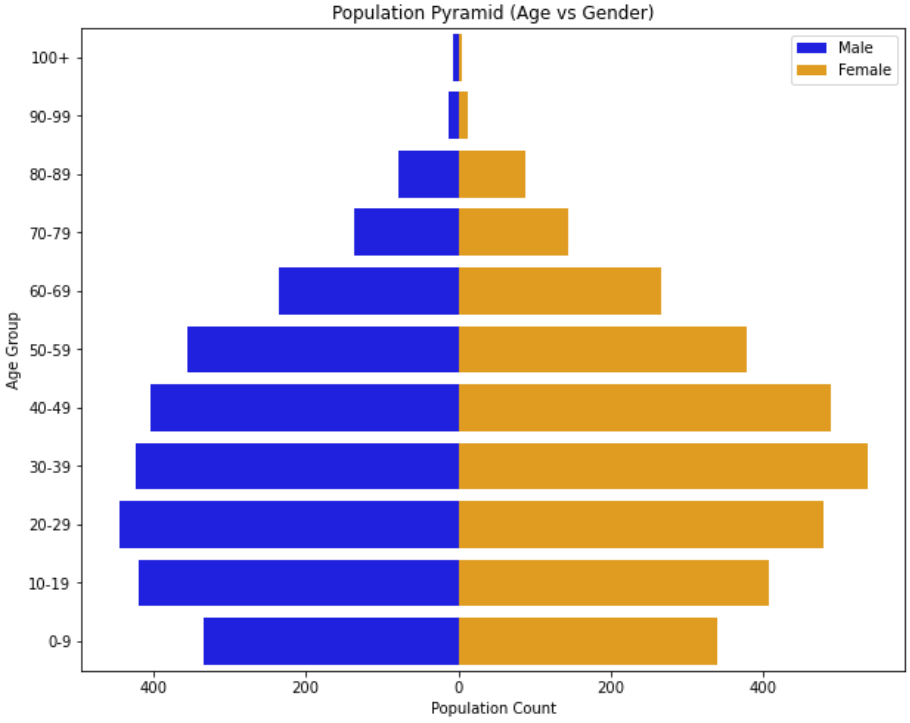
Infirmity

Value	Count	Frequency (%)
None	5958	99.4%
Physical Disability	10	0.2%
Deaf	9	0.2%
Disabled	7	0.1%
Unknown Infection	5	0.1%
Mental Disability	4	0.1%
Blind	2	< 0.1%

Majority of the population do not have any infirmity.

Detailed Analysis

Population pyramid



The distribution of age and gender in the chart reveals a larger proportion of young adults and middle-aged individuals, who are typically part of the workforce, in comparison to the number of children (aged 0-9) and elderly population (aged between 70 to 100+), which are significantly lower. The shape of the population pyramid indicates that the population is decreasing since the base of the pyramid is narrower than the top, which suggests a low birth rate. Moreover, the small number of elderly people in the population may indicate that the town has a high life expectancy, which results in a low death rate.

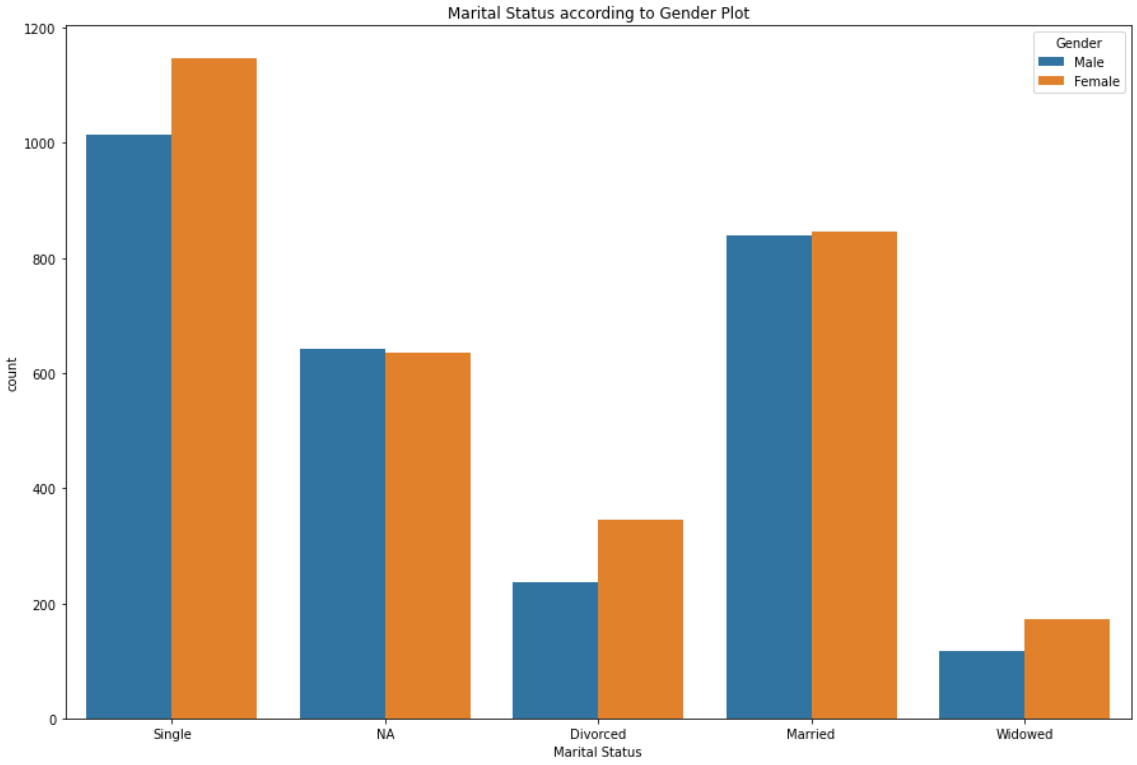
Divorce and Marriage Rate

Marital Status	count	mean	std	min	25%	50%	75%	max
Divorced	581.0	42.612737	17.970084	18.0	27.0	39.0	55.00	105.0
Married	1684.0	50.267815	16.910484	16.0	38.0	49.0	62.00	105.0
NA	1279.0	8.894449	5.151899	0.0	4.5	9.0	14.00	17.0
Single	2161.0	37.144840	14.057515	18.0	25.0	35.0	47.00	94.0
Widowed	290.0	67.827586	18.325147	18.0	66.0	71.0	78.75	104.0

Based on the descriptive analysis, it was observed that the number of married individuals in the population is higher than the number of divorced individuals. Specifically, 28.1% of the population are married while 9.7% are divorced. To calculate the crude marriage rate per thousand, the total count of married individuals is divided by two and then divided by the total population count, which is then multiplied by 1000. The same approach is used to calculate the male and female divorced rates. The marriage rate of 140 per thousand means that for every 1000 people in the population, 140 are married. This is calculated by dividing the total number of married people by the total population, multiplying by 1000, and then dividing by 2 (since each marriage involves two people).

Similarly, the female divorced rate of 58 per thousand means that for every 1000 females in the population, 58 are divorced. To obtain this figure, the total number of females who are divorced is divided by the total female population, and then the result is multiplied by 1000.

The male divorced rate of 39 per thousand means that for every 1000 males in the population, 39 are divorced. To obtain this figure, the total number of males who are divorced is divided by the total male population, and then the result is multiplied by 1000.



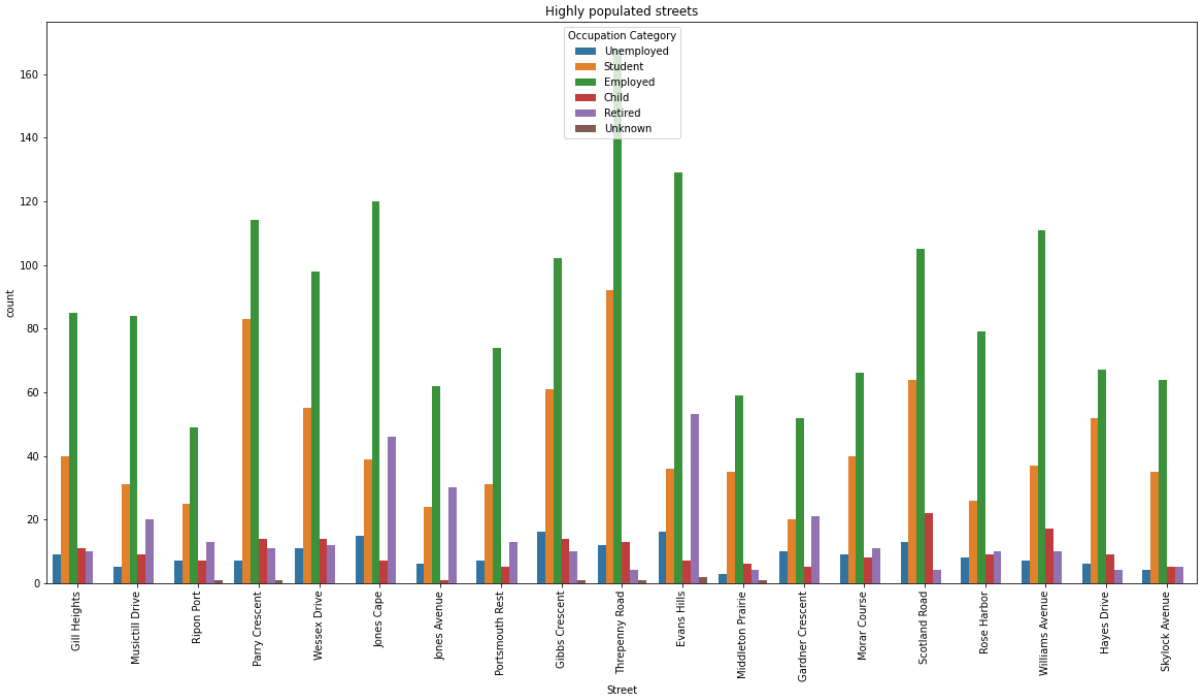
Birth and Death Rate

From the birth rate calculation, it can be observed that the town is not growing. The crude birth rate was calculated using Crude Birth Rate = (Number of births/Total Population) * 1000 (Murray, 1997). To calculate the death rate, deaths are estimated by comparing the number of people in different age groups, specifically those between the ages of 80 and 100+. The crude birth rate is the number of live births that occur during a specific time period per 1,000 people in a particular population during a specific period of time. In this case, the calculation indicates that there were 8 live births per 1,000 people in the population.

The crude death rate, on the other hand, refers to the number of deaths per 1,000 people in a given population during a specific time period. In this case, the calculation indicates that there were 3 deaths per 1,000 people in the population.

Housing and Occupancy

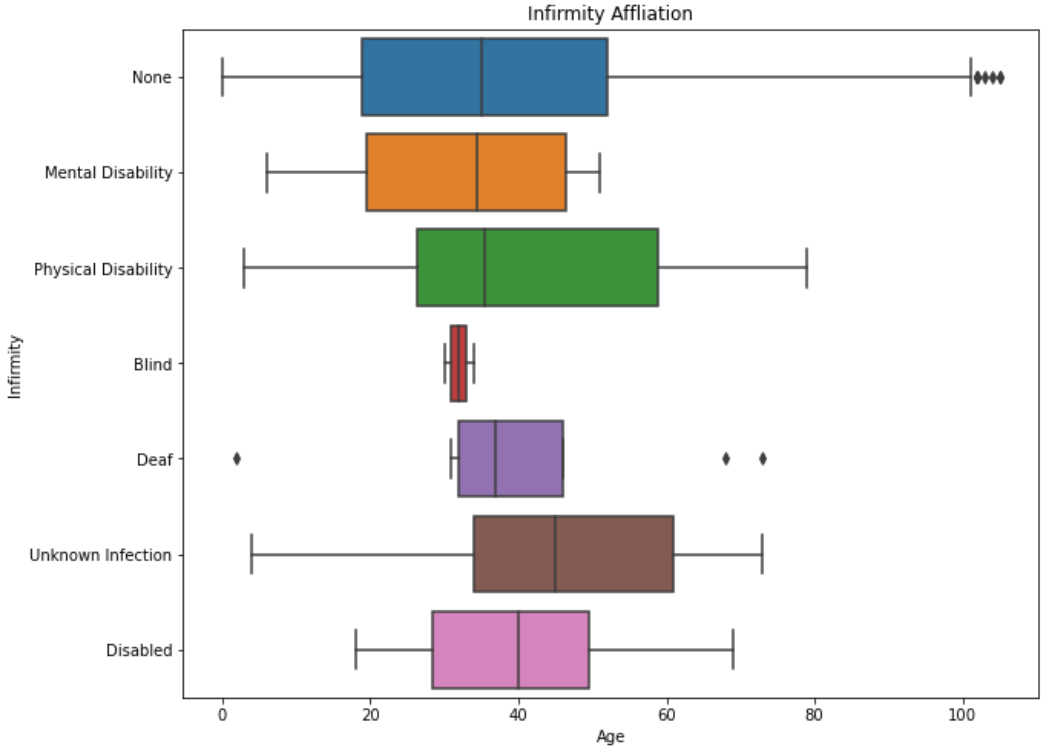
The town has 105 streets and 2328 houses, with different levels of population density. On average, there are 4 occupants per house. Out of the 105 streets, 19 have a high population density and are home to 1114 houses and majority of the houses are occupied by the employed population, while the remaining 86 streets have a lower population density and contain 1214 houses. Additionally, it is estimated that there are 189 lodgers residing with families who have children, and there may be even more lodgers in other family homes.

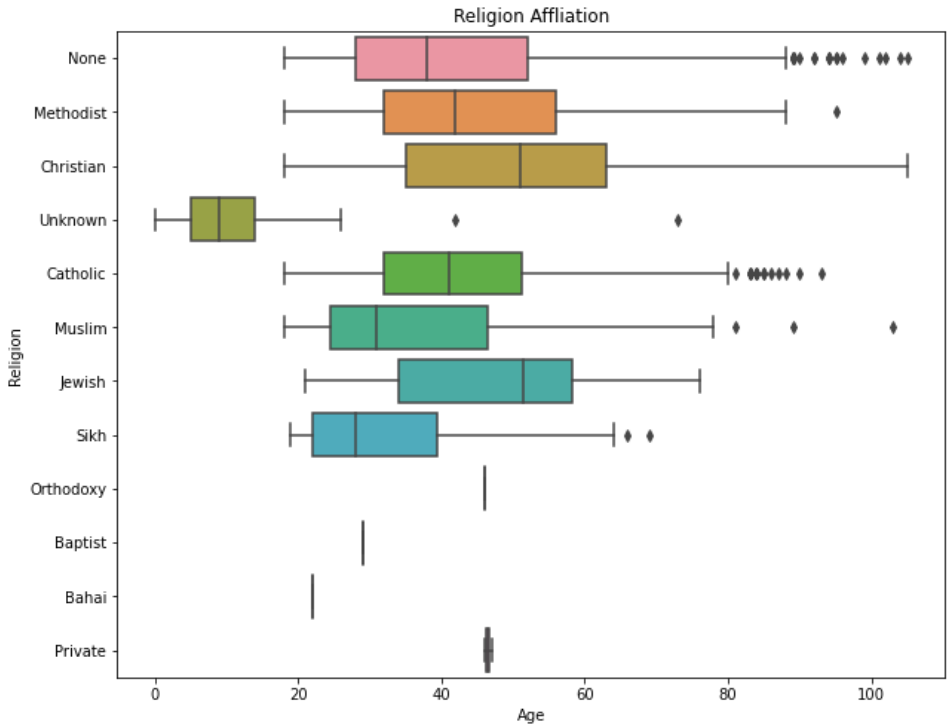


Religion and Infirmity

A large proportion of the population in the town does not have any religious affiliation or infirmity record. This indicates that the community is neutral in terms of religious practices and relatively healthy, as there is no evidence of any life-threatening illnesses or epidemics. Although Christianity remains the primary religion in the town and there is already a church to serve its followers. While it is possible that other religions may gain more followers in the future, the current number of non-Christian religious adherents is not significant enough to justify the construction of a new religious institution over other pressing needs of the population. Given that religion is a

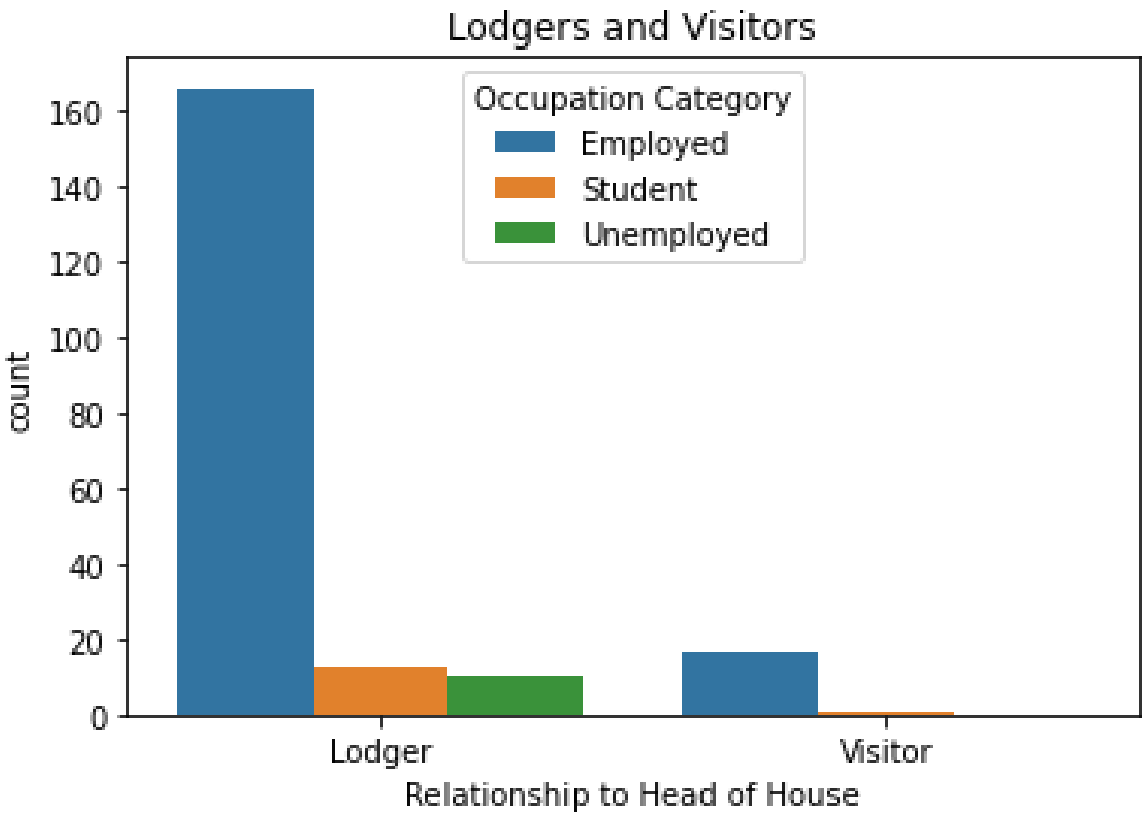
personal choice and a majority of the population does not have a particular affiliation, no further analysis is needed in that regard. Similarly, since there are no significant health concerns, there is no need to investigate further.





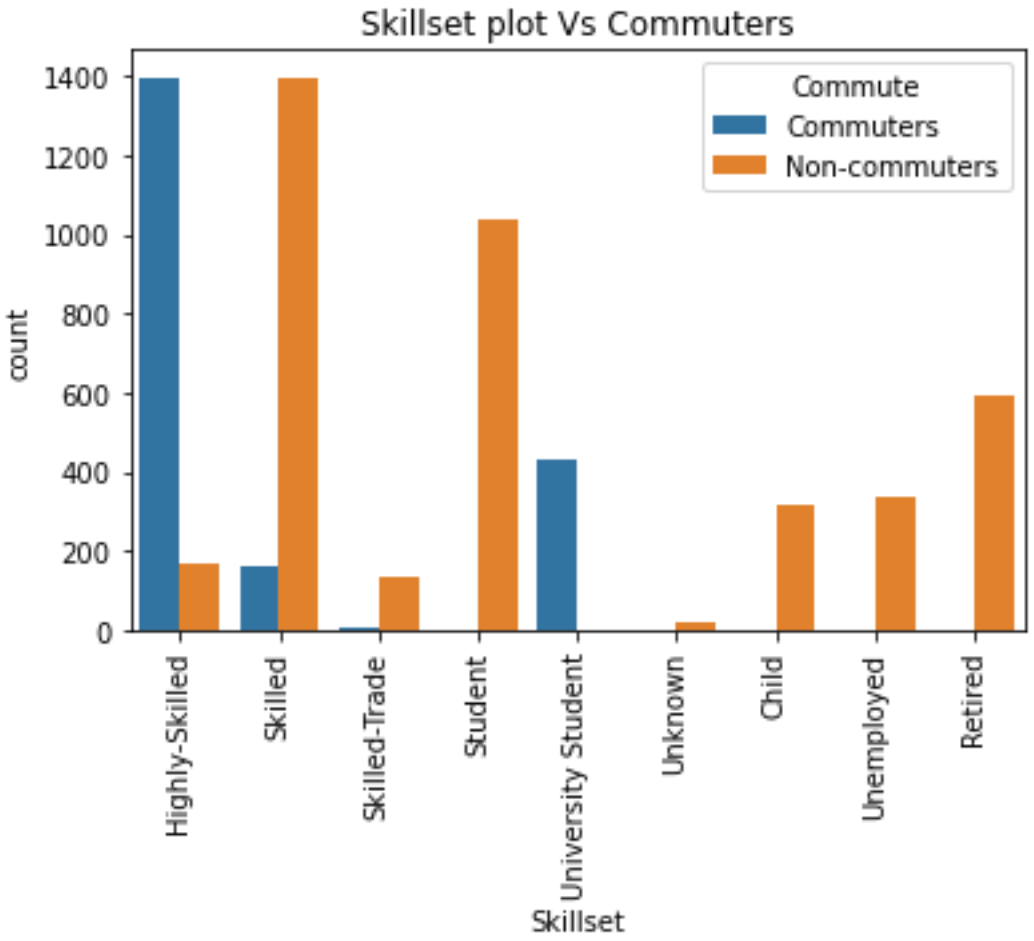
Migration

The census data identifies two categories of individuals who can be classified as immigrants to the town: lodgers and visitors. The total number of lodgers and visitors in the town is 207. Approximately 21 % of the lodgers are divorced, 76% are Single and the rest are widowed. 78% of the visitors are single and 22% of the visitors are divorced. Over 50% of the people who live in lodgings are either highly skilled workers or university students. Skilled workers tend to stay for a long time and make a positive contribution to the town, while university students may stay for a shorter, more temporary period of time. According to the data, there are 35 immigrants per thousand people in the town, which means that out of every 1,000 people in the town, 35 of them are immigrants.



Employment and Commuters

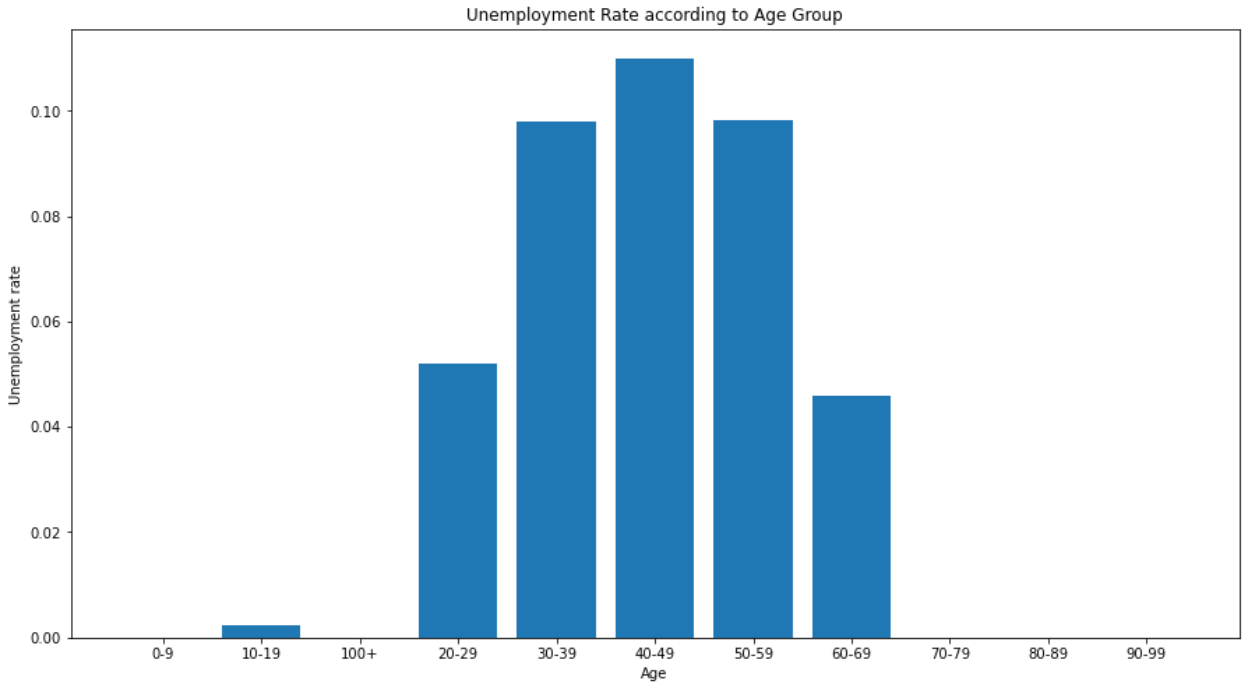
To group occupations by commute needs based on skillset and geographic location, it was found that there are 4,000 non-commuters in the town compared to 1,995 commuters. The reason for this is that proficient employees usually look for employment prospects in the metropolitan areas surrounding the town, while young adults attend universities in those cities. Those who don't commute include various groups such as students who attend local schools, retired individuals, administrative, technicians and civil service workers, among others. Also, the town needs to expand and create more job opportunities to attract commuters and promote population growth. The population of the town includes a total of 433 university students who are among the top commuters, which when compared to the total population, results in a proportion of 72 university students per thousand people residing in the town.



Unemployment Rate

The unemployment rate in the town is relatively high at 5.6% (ONS, 2020). This can be attributed to the fact that many employment opportunities are located outside of the town. Further analysis shows that 1.1% of the population aged between 40-49 are unemployed, 0.98% of the population aged between 50-59 are unemployed, and 0.978% of the population aged between 30-39 are unemployed.

.



RECOMMENDATION

Constructing a train station in the town can facilitate its growth by enabling easy connections with neighboring cities, attracting foreign investment, and promoting economic development. In addition, this infrastructure development can create employment opportunities and reduce the high unemployment rate in the town, while increasing its internally generated revenue for investment in other essential projects. Although it's important to invest in infrastructure projects like housing, expansion of schools, and skills development programs, they may not be the top priority right now,

considering the town's high unemployment and birth rates. Instead, it's recommended to focus on investing in elderly care for the future, as the working population is expected to retire soon.

REFERENCES

Government Digital Service (2016). *Working after State Pension age*. GOV.UK. Available at: https://www.gov.uk/working-retirement-pension-age. [Accessed 12/04/2023]

Murray, B.G., 1997. On calculating birth and death rates. *Oikos*, pp.384-387.

NSPCC (2022). *Children and the law | NSPCC Learning*. NSPCC. Available at: https://learning.nspcc.org.uk/child-protection-system/children-the-law. [Accessed 12/04/2023]

ONS (2021). *National life tables – life expectancy in the UK - Office for National Statistics*. Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/lifeexpectancies/bulletins/nationallifetablesunitedkingdom/2018to2020. [Accessed 10/04/2023]

www.ons.gov.uk. (n.d.). *SOC 2020 Volume 1: structure and descriptions of unit groups - Office for National Statistics*. Available at: https://www.ons.gov.uk/methodology/classificationsandstandards/standardoccupationalclassificationsoc/soc2020/soc2020volume1structureanddescriptionsofunitgroups. [Accessed 12/04/2023]

www.psa.gov.qa. (n.d.). *What is Census?* Available at: https://www.psa.gov.qa/en/statistics1/StatisticsSite/Census/census2010/pages/whatiscensus.aspx. [Accessed 20/04/2023]