

CENSUS DATA REPORT

This report compares individuals within a small town connected by motorways to two or more larger cities. It provides precise population numbers to the government for better planning and investment decisions for an unoccupied piece of land. However, the dataset required cleaning due to missing values and incorrect data entries before analyses can be done.

DATA CLEANING

An overview of issues to address is shown below, as highlighted in the Jupyter Notebook.

- Presence of a duplicate sample
- Missing values in Marital Status and Religion columns
- Inconsistent and blank inputs in the Religion column
- Blank inputs in Infirmary, Occupation, Age, Surname, and First Name columns
- Misspelling and blank inputs in the Relationship to Head of House column
- Inconsistent and wrong datatype in the Age column
- Renaming of some columns by replacing spaces with underscore

The duplicate sample was identified and removed from the dataset.

The missing values (NaN) in the marital status were traced to individuals under 18. In the UK, a minor is someone under the age of 18. Hence, all were grouped as Minors. Furthermore, married children under 18 were changed to single (Ministry of Justice, 2022). For divorced individuals, you need to have been married for at least a year before filing a case; hence, individuals that were 18 were not possible legally; therefore, they were replaced as single (Government Digital Service, 2011).

Two unique cases of underaged single moms with no parents or traces of being married were left in the dataset alongside their kids since they are allowed to move out between 16 and 18 (NSPCC, 2021). A seventeen-year-old female head of the house alongside her husband was changed to singles.

Infirmary column entries with blank inputs were replaced with Unknown since the individuals chose not to disclose their status.

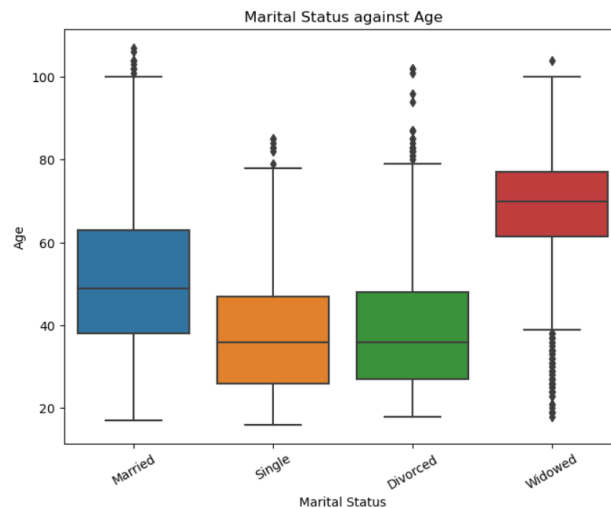
An investigation was done using an individual's age range to fix blank inputs in occupation. Empty inputs in first_name, and surname columns were replaced with Unspecified where no similarities exist after a demographic search using house_number, surname and street.

The relationship to the head of house had niece wrongly spelt and fixed while there was a blank input. Information was obtained using house_number, street, and surname before replacing with daughter after further inspection.

One blank input in age was investigated using surname and street before replacing it with the median age based on her occupation. The age data type was also changed from an object to an integer.

Missing values (NaN) in the religion column were addressed by replacing them with the religion of the head of the house based on their surname, house_number, and street. Eight individuals with private, undecided, or blank as religion, were assigned unknown and nope replaced with none. 444 individuals, including 18 and above, with no head of the house, were replaced with unknown as no one can force a child or young person to adopt or to stop following a religion (UNCRC, 2020).

Although there is no longer forced retirement age of 65, individuals (65 and older) who were unemployed were replaced as retired (GOV.UK, 2016).



The plot demonstrated the existence of outliers across, as was expected. Individuals above 75 can be divorced, married, single, or widowed, with the majority as married and divorced. Individuals who were divorced and 18 or married and below 18 were referred to as outright lies, and were replaced as single.

POPULATION DATA

After extensively cleaning the dataset, new features were added to aid in analyses and visualizations.

```
Data columns (total 14 columns):
#   Column                                     Non-Null Count  Dtype
---  -
0   House_Number                             9658 non-null   int64
1   Street                                    9658 non-null   object
2   First_Name                               9658 non-null   object
3   Surname                                   9658 non-null   object
4   Age                                        9658 non-null   int32
5   Relationship_to_Head_of_House            9658 non-null   object
6   Marital_Status                           9658 non-null   object
7   Gender                                    9658 non-null   object
8   Occupation                                9658 non-null   object
9   Infirmary                                9658 non-null   object
10  Religion                                  9658 non-null   object
11  Occupation_Category                       9658 non-null   object
12  Household                                  9658 non-null   int64
13  Age_Group                                 9658 non-null   category
dtypes: category(1), int32(1), int64(2), object(10)
memory usage: 1.3+ MB
```

Occupations were grouped into employed, unemployed, students, retired and child to create Occupation_Category. All students, including PhD and university students, were categorized as students, while those with all job types were grouped as employed.

Age was grouped into a band of 5 for the population pyramid's Age_Group.

A person living alone or a group of people not necessarily related residing at the same address is called a Household. It can consist of a single family, more than one family or no family in the case of a group of unrelated people (Sharfman, 2022). Hence, the household was created by grouping individuals using house number and street.

	House_Number	Age	Household
count	9658.000000	9658.000000	9658.000000
mean	37.434044	34.908159	4.087803
std	41.827855	21.401611	2.703526
min	1.000000	0.000000	1.000000
25%	9.000000	17.000000	3.000000
50%	22.000000	34.000000	4.000000
75%	47.000000	50.000000	5.000000
max	214.000000	107.000000	22.000000

According to the statistical information, the town has a wide spread of individuals of varying ages. The 75th percentile implies that $\frac{3}{4}$ of the population is 50 years and below, further corroborated by an age pyramid.

Marital_Status

Value	Count	Frequency (%)
Single	3411	35.3%
Married	2652	27.5%
Minor	2434	25.2%
Divorced	847	8.8%
Widowed	314	3.3%

Gender

Value	Count	Frequency (%)
Female	5041	52.2%
Male	4617	47.8%

Household – Top 5

Value	Count	Frequency (%)
4	2788	28.9%
5	1760	18.2%
2	1664	17.2%
3	1554	16.1%
1	672	7.0%

Occupation_Category

Value	Count	Frequency (%)
Employed	5166	53.5%
Student	2574	26.7%
Retired	776	8.0%
Child	595	6.2%
Unemployed	547	5.7%

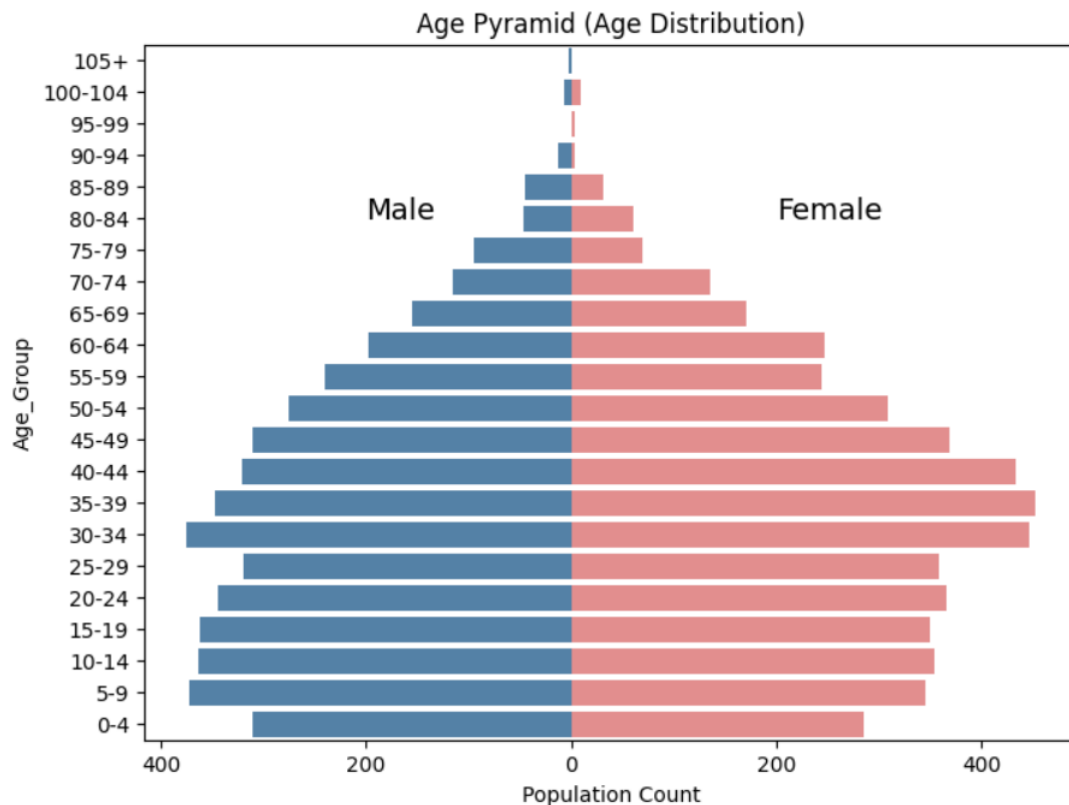
Religion – Top 4

Value	Count	Frequency (%)
None	3968	41.1%
Christian	2794	28.9%
Catholic	1354	14.0%
Methodist	826	8.6%

EXPLORATORY ANALYSIS

Age Distribution

According to the age pyramid below, the shape had a bulge in the middle, indicating the town's population were primarily middle-aged persons aged 30 to 44. This suggests that in the future, there will be more middle-aged persons (if the retirement age is 65) than retired, school-aged children and young people. 38.64 percent of the town's population is aged 25 to 49.



Furthermore, a contracting bottom and a narrow top in the pyramid indicate an aging population resulting from long life-expectancy, low birth and death rates. In general, the town showed a reproductive aging population filled with more adults between the ages of 25 and 49 than young and old aged individuals.

Further analysis shows that 52.2 percent of the total population were females, with over ½ of the people comprising those aged 35 and below.

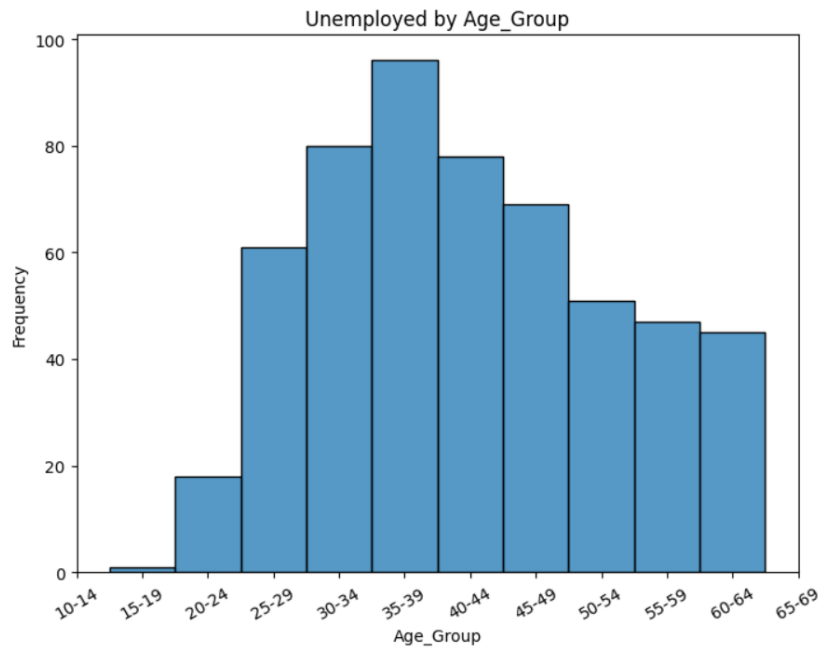
	count	mean	std	min	25%	50%	75%	max
Gender								
Female	5041.0	35.378298	20.896407	0.0	19.0	35.0	50.0	104.0
Male	4617.0	34.394845	21.930691	0.0	16.0	33.0	50.0	107.0

Unemployment Trend

53.5 percent of the population was employed, with females accounting for 53.19 percent. Unemployment accounted for 9.57 percent of the workforce, and this is due to the unavailability of jobs in the city hence, the rise in commuters. Fifty percent of employed and unemployed individuals were 41 years and below.

	count	mean	std	min	25%	50%	75%	max
Occupation_Category								
Child	595.0	1.981513	1.394311	0.0	1.0	2.0	3.0	4.0
Employed	5166.0	42.382695	12.170843	19.0	32.0	41.0	52.0	67.0
Retired	776.0	76.418814	7.639531	65.0	71.0	74.0	81.0	107.0
Student	2574.0	13.544678	5.307723	5.0	9.0	14.0	18.0	31.0
Unemployed	547.0	41.773309	11.145712	16.0	33.0	41.0	50.0	64.0

There were more unemployed individuals aged 35 to 39 than any other age group in the town. In summary, individuals aged 30 to 44 accounted for 46.44 percent of the unemployed workforce, as seen below.



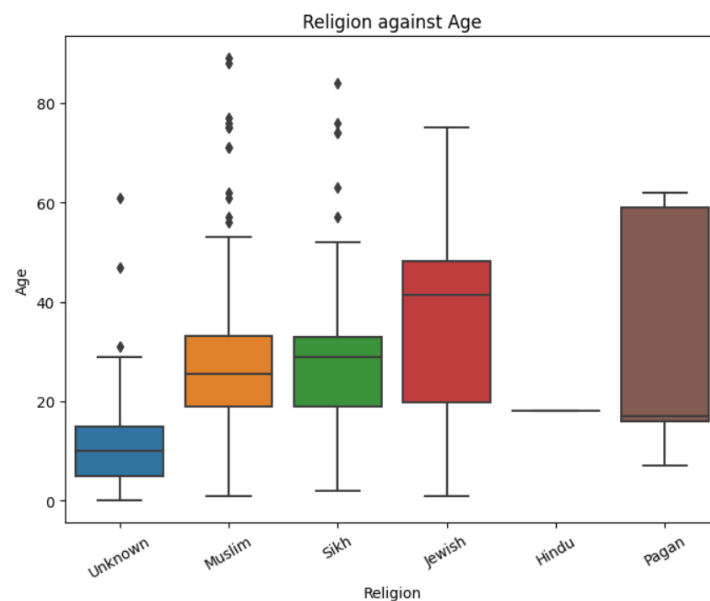
Religious Affiliation

Catholic and Methodist accounted for 22.6 percent of the population, with 50 percent aged 34 and below; this justifies the existence of a place of worship for Catholics. Despite Sikh and Muslim religions accounting for less than 3 percent of the town's population, it was seen that younger people practiced the religion more than older people.

	count	mean	std	min	25%	50%	75%	max
Religion								
Catholic	1354.0	32.173560	19.669633	0.0	15.00	33.0	46.00	104.0
Christian	2794.0	40.755906	22.622935	0.0	21.00	41.0	59.00	104.0
Hindu	1.0	18.000000	NaN	18.0	18.00	18.0	18.00	18.0
Jewish	30.0	36.633333	18.159226	1.0	19.75	41.5	48.25	75.0
Methodist	826.0	33.858354	20.975149	0.0	16.00	34.0	48.00	103.0
Muslim	156.0	27.602564	18.003111	1.0	19.00	25.5	33.25	89.0
None	3968.0	35.173135	20.096792	0.0	21.00	34.0	48.00	107.0
Pagan	5.0	32.200000	26.147658	7.0	16.00	17.0	59.00	62.0
Sikh	77.0	28.363636	17.490531	2.0	19.00	29.0	33.00	84.0
Unknown	447.0	9.856823	6.680545	0.0	5.00	10.0	15.00	61.0

Individuals of no religion made up 41.1 percent of the town's population, with 50 percent aged 34 and below, while Christians accounted for 28.9 percent of the people, with 50 percent aged 41 and below. Hence, this suggests that more individuals in the future are less likely to practice

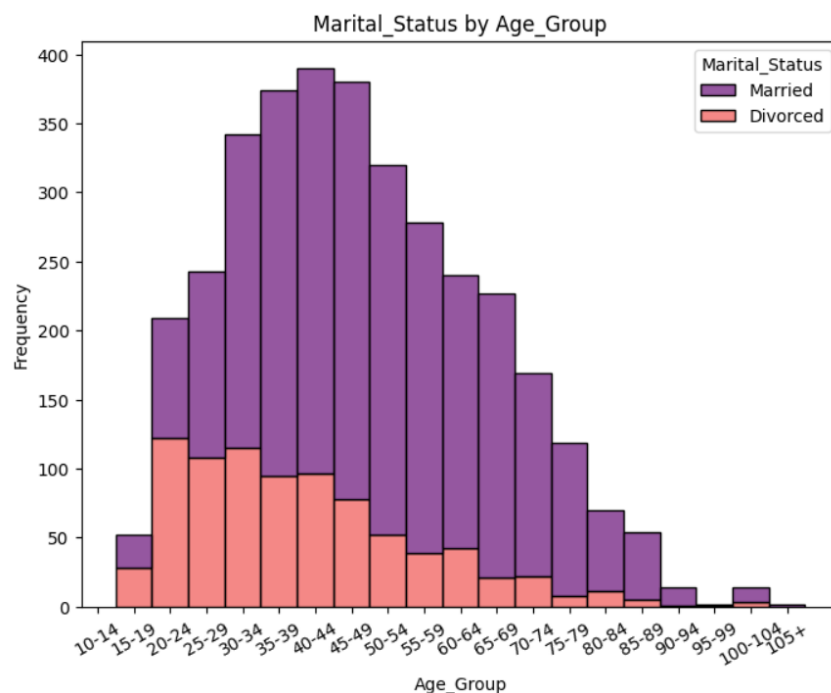
any religion and a wide spread of Christians with varying ages; therefore, there is no need for an additional religious building. A standard deviation of 22.62 and a mean of 40.76 showed that 63.21 percent of Christians were aged 18 and 64. 5 percent of the population had unknown religion or other religions.



Less than 0.8 percent of the population had an infirmity, and 8 percent were retired. Even in the future, there will be more middle-aged people, which justifies why there would be no need for an emergency medical building or old age care.

Divorce and Marriage Rate

36.73 percent of adults who had reached marriageable age were married, with 50 percent aged 49 and below. The standard deviation indicated the vast age range in which people were married, ranging from young to old, and there were more married women than men. There were more married individuals aged 45 to 49 than any other age group in the town.

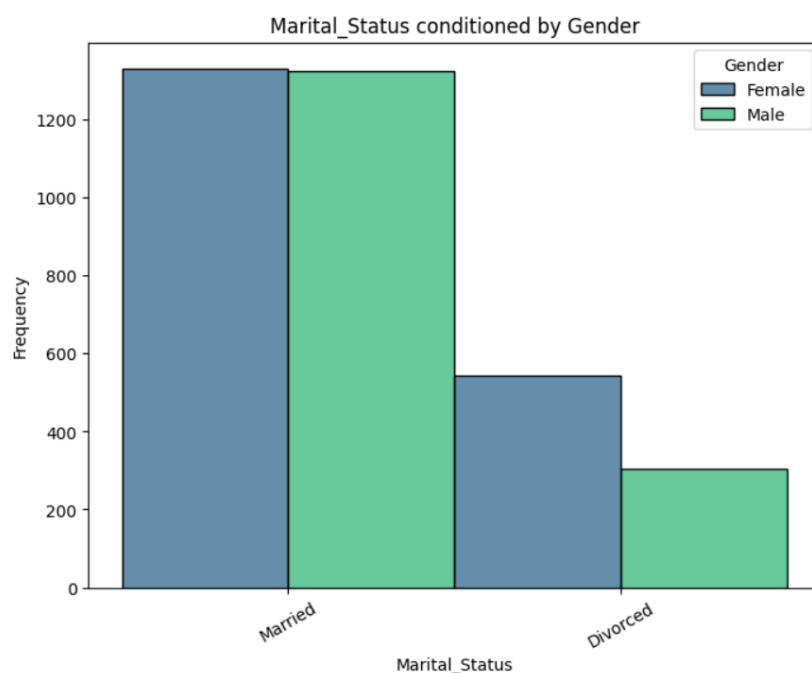


11.97 percent of adults who had reached divorceable age were divorced, with 50 percent aged 37 and below. The standard deviation showed that the risk of a divorce occurring at a young age was higher than at a later age, with over ¾ of the cases being 49 and below. Women accounted for 63.99 percent of individuals that were divorced.

	count	mean	std	min	25%	50%	75%	max
Marital_Status								
Divorced	847.0	40.073200	16.099112	19.0	28.0	37.0	49.0	102.0
Married	2652.0	50.792232	16.731132	18.0	38.0	49.0	63.0	107.0
Minor	2434.0	8.725555	5.067837	0.0	5.0	9.0	13.0	17.0
Single	3411.0	37.245089	13.570292	16.0	26.0	36.0	47.0	85.0
Widowed	314.0	64.391720	19.540146	19.0	62.0	70.0	77.0	104.0

The crude divorce rate was estimated by referring to the town's total number of married people, i.e., those currently married, divorced, and widowed; thus, 44.43 per thousand married adults per year over the past five years were divorced. Similarly, the crude marriage rate was estimated to be 139.10 per thousand married adults per year over the past five years who were still married. 39.48 percent of the total population can be classified as married adults.

$$\text{Marriage Rate} = \frac{\text{No of Married per year}}{\text{Married Population}} \quad \text{Divorce Rate} = \frac{\text{No of Divorced per year}}{\text{Married Population}}$$



Occupancy Level

The 50th percentile (or median) for households implies that over ½ of the houses in the city were occupied by four or fewer persons. Assuming a house must have four persons to be sufficiently occupied, 514 homes would be classified as overuse, as seen below.

Household	1	2	3	4	5	6	7	8	9	10	11	12	13	14	16	17	20	21	22
No_of_Individuals	672	1664	1554	2788	1760	660	105	16	54	60	110	36	13	28	16	17	40	21	44
No_of_Houses	672	832	518	697	352	110	15	2	6	6	10	3	1	2	1	1	2	1	2

Divorcees occupied 9.53 percent of overused homes, with women accounting for 60.21 percent. Following the divorce, these women either gained ownership of the house or took sole responsibility for paying the rent. Moreso, singles occupied 37.85 percent of the overused homes, and this is due to the town's high influx of migrants and students.

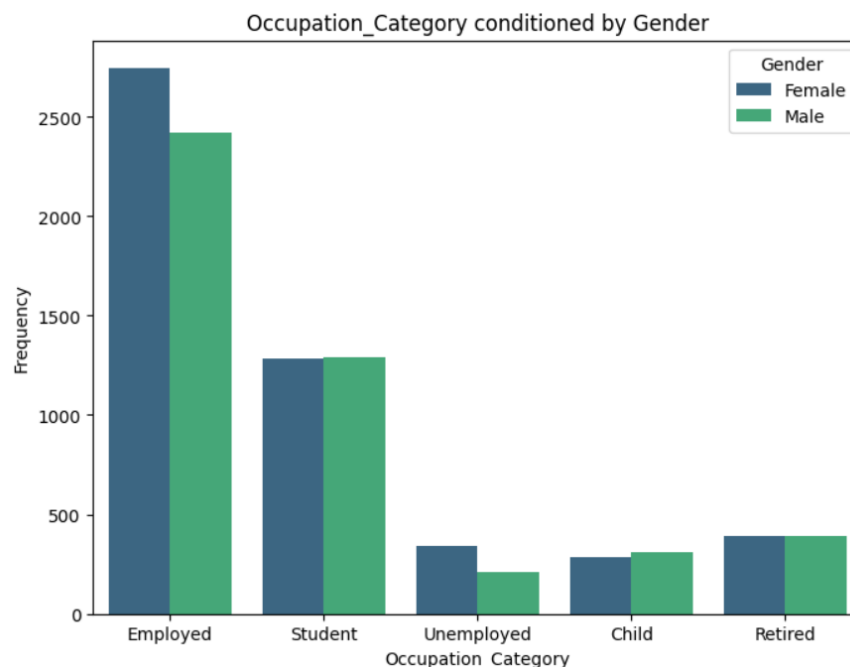
Household	1	2	3	4	5	6	7	8	9	10	11	12	13	14	16	17	20	21	22
count	672	1664	1554	2788	1760	660	105	16	54	60	110	36	13	28	16	17	40	21	44
unique	1	17	19	21	18	16	10	5	9	9	7	4	6	5	4	6	6	5	7
top	Head	Head	Head	Head	Son	Son	Son	None	Daughter	Daughter	None	Daughter	Lodger	None	Lodger	Daughter	Lodger	Daughter	Lodger
freq	672	832	518	697	407	179	36	7	16	14	30	17	7	12	9	8	15	9	19

Lodgers make up highly overused homes. These individuals must have migrated with their families, which explains why so many of them were in crowded homes. Overused houses comprise 15.9 percent of the total number of houses in the town, i.e., 84 percent are either sufficiently occupied or under-occupied, and they make up 69.2 percent of the town's population. 62.54 percent of the houses in the town need to be sufficiently occupied, which justifies why there is no need to build new houses.

Commuters and University Students

Students who were commuters included PhD and university students, while commuters were employers whose jobs required commuting. The number of commuting employers was estimated using the assumption that writers, primary and secondary school teachers, retailers, baristas, and community and local government workers were non-commuters.

56.76 percent of the town's population were commuters, with students accounting for 6.39 percent. Migrants make up 6.95 percent of the total commuters, including students and employers.



The plot above depicts the effect of commuters on the town. 88.74 percent of commuters were employed. This is expected for a city with a high proportion of middle-aged individuals, closely followed by young individuals, because, in today's world, individuals of this age range make up employers and students.

Birth and Death Rate

The number of births per year was estimated by averaging the number of individuals between 0 and 4. Logically, it makes sense to infer that these individuals were born within the past five years.

$$\begin{array}{lcl} \text{Crude Birth Rate} & & \text{Crude Death Rate} \\ = \frac{\text{No of Birth per year}}{\text{Overall Population}} & & = \frac{\text{No of Death per year}}{\text{Overall Population}} \end{array}$$

The mean value obtained was used to estimate the yearly crude rate as 12.32 for every thousand individuals in the town. From this, it can be seen that more migrants tend to move into the city than citizens giving birth.

Assuming the life expectancy is 65 years, the number of deaths per year was obtained by estimating the difference in the age group from age 65 and dividing by the band length. As individuals get to the age of retirement, their life expectancy reduces; hence, the reason for the assumption.

The value obtained was used to calculate the yearly crude rate as 6.69 for every thousand individuals in the town. In 10 years, there would be an estimated 12.89 percent of old-aged people and 16.83 percent of school-aged children of the current population.

Recommendations

Having a town filled with middle-aged and young people is evident in the number of individuals who are commuters. Middle-aged individuals make up the workforce in today's world hence, the reason for the high percentage of commuters. Moreso, the absence of a university in the town gives room for more commuters, i.e., university students.

Furthermore, the increasing number of migrants who make up the middle-aged and young people justifies the need for a train station in the unoccupied plot of land over other development. The presence of a train station will not only increase the number of migrants but will aid economic growth as more investors will be more than willing to invest or create business opportunities in the town, thus, reducing the unemployment rate.

A town with 56.76 percent of its population as commuters and 9.57 percent as unemployed should prioritize creating faster and easier means of transportation for its citizens to aid economic growth since these categories of people make up the taxpayers and, in addition, would create job opportunities. In the UK, compared with 3.8 percent, this unemployment rate exceeds the highest rate ever gotten between December 2022 and February 2023 (ons.gov.uk, 2021). Training would have been essential, but it's not a priority over school children with a higher population.

As initially stated, there are more underused homes in the town than overused ones, so building more houses for migrants should not be an option. 9.53 percent of underused houses are made up of divorcees, with women accounting for 60.21 percent. Making these homes available for migrants through renting can help reduce overcrowding in highly overused homes.

The only religion with many followers, mainly aged 18 and 64, is the Christian religion, which shows excellent growth potential and may require a building; building a train station takes precedence over any other building considering the percentage of commuters. For a town with a catholic church, which shows excellent prospects based on the high rate of young people, there is evidence of a growing population in the future. More reason to justify why building a

religious home should be overlooked is that 41.1 percent of individuals in the town belong to no religion.

For a community with 9.96 percent of the population above 65 years of age, less than 0.8 percent with infirmity, and a low birth rate, investing in an emergency medical building or old aged care look like a waste of resources with little or no dividend in the future. The bulgy middle showed a reproductive aging population resulting from middle-aged people and migrants who chose to bring their families with them, as seen from the age pyramid.

With 26.65 percent of the total population as students and 77.12 percent aged 18 and below, it is inevitable to overlook the importance of increasing school spending. In addition, it's projected that in ten years, the percentage of school-aged children will be 16.83 and 12.89 for old-aged people. Aside from citizens, migrants also comprise a reasonable number of these individuals. With their contributions to the city, it's paramount for their children to benefit from increased spending on schooling. Furthermore, if additional funds are left, the government may want to look into old aged care considering the aging population.

In conclusion, a train station and investing in school spending will aid ease of commuting and promote economic and population growth through increased migration.

References

- Ministry of Justice (2022). *Implementation of the Marriage and Civil Partnership (Minimum Age) Act 2022*. [online] GOV.UK. Available at: <https://www.gov.uk/government/news/implementation-of-the-marriage-and-civil-partnership-minimum-age-act-2022> [Accessed 19 Apr. 2023].
- GOV.UK. (2016). *Check your State Pension age*. [online] Available at: <https://www.gov.uk/state-pension-age> [Accessed 19 Apr. 2023].
- Sharfman, A. (2022). *Families and households statistics explained*. [online] ons.gov.uk. Available at: <https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/families/articles/familiesandhouseholdsstaticsexplained/2021-03-02#:~:text=called%20adult%20children,-.Household,sitting%20room%20or%20dining%20area> [Accessed 19 Apr. 2023].
- The Children and Young People's Commissioner Scotland. (2020). Article 14 – *I have the right to have my own thoughts and beliefs and to choose my religion with my parents' guidance*. [online] Available at: <https://www.cypcs.org.uk/rights/uncrc/articles/article-14/#:~:text=Article%2014%20of%20the%20UNCRC,to%20adopt%20a%20religion%2C%20and> [Accessed 19 Apr. 2023].
- Government Digital Service (2011). *Get a divorce*. [online] GOV.UK. Available at: <https://www.gov.uk/divorce> [Accessed 19 Apr. 2023].
- Ons.gov.uk. (2021). *Unemployment – Office for National Statistics*. [online] Available at: <https://www.ons.gov.uk/employmentandlabourmarket/peoplenotinwork/unemployment> [Accessed 19 Apr. 2023].
- NSPCC (2021). *Moving out*. [online] NSPCC. Available at: <https://www.nspcc.org.uk/keeping-children-safe/in-the-home/moving-out/> [Accessed 25 Apr. 2023].