

CENSUS DATA PROJECT

An Analysis of a Town Population Data

INTRODUCTION

The purpose of this project is to conduct an analysis of the population data for a moderately sized town situated between two cities. The primary objective is to provide recommendations for the town regarding the best use of an unutilized plot of land. The town requires expert guidance on what type of development would be suitable for the vacant land and what investment strategy would be most beneficial.

The ensuing sections will outline the comprehensive process for data collection and detailed analysis. The analysis will focus on key factors such as birth rate, death rate, prospective religion for children, and occupancy level, which will aid in generating recommendations for the town.

DATA CLEANING

The data cleaning process involved handling data types and missing values. The cleaning steps taken are outlined below:

Age:

An empty age row in the data was identified as belonging to a university student. The age was replaced with 20, which is the median age range between 18-22, where most university students fall.

Marital Status:

A considerable number of missing values were found in the marital status column, particularly for individuals under 18 years old. Since the legal age for marriage in the UK is 18 years (GOV.UK, 2023), null values were replaced with 'Not Applicable' for these individuals. A single blank marital status column was identified as belonging to a son in a household, and 'Single' was assigned to it.

Religion:

Missing values in the religion column for individuals under 18 and those who inputted Jedi as their religion were assigned 'Undecided.' This is because Jedi is a fictional religion (Wikipedia, 2023), and it was assumed that the person was undecided and decided to apply fiction. For those who inputted 'housekeeper' as their religion, 'None' was assigned, as it was deemed a mistake.

Occupation:

Empty values in the occupation column were filtered, and a value was assigned after examining other attributes of the subject rows. One row had an empty occupation column, and since the person was over 70 years old, 'Retired' was assigned to it since most rows with this age range were retired. The retirement age of the population was also calculated to provide further explanation.

Gender:

Three empty values were found in the gender column. The rows were inspected, and the first names were checked to determine if they were masculine or feminine (babynames.com). All the names were feminine, so 'Female' was assigned to the three rows.

Relationship to Head of House:

Two missing values were found in the 'Relationship to head of house' column. Rows of all members of the household they were in were filtered, and it was observed that they were sons of lodgers. They had similar surnames to older members of the household, and they were under 18 years old, so they were assigned 'Lodger'.

Infirmity:

Missing values in the infirmity column were assumed to be 'None'. This is based on the assumption that since the individual had no infirmity, nothing was inputted.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8377 entries, 0 to 8376
Data columns (total 14 columns):
#   Column                                     Non-Null Count  Dtype
---  -
0   House Number                             8377 non-null   object
1   Street                                   8377 non-null   object
2   First Name                               8377 non-null   object
3   Surname                                  8377 non-null   object
4   Age                                       8377 non-null   int64
5   Relationship to Head of House            8377 non-null   object
6   Marital Status                           8377 non-null   category
7   Gender                                   8377 non-null   category
8   Occupation                               8377 non-null   object
9   Infirmity                                8377 non-null   object
10  Religion                                  8377 non-null   object
11  Age Group                                8363 non-null   category
12  Employment Status                        8377 non-null   object
13  Prospective Religion and Current Religion 8377 non-null   object
dtypes: category(3), int64(1), object(10)
memory usage: 745.3+ KB
```

Fig: Information about the clean census DataFrame

EXAMINING THE AGE DISTRIBUTION

To conduct a more comprehensive analysis, a new column entitled "Age Group" was introduced. The bands were carefully selected to differentiate the minimum legal age (gov.uk) for working full time from the younger population. While most groups had a 10-

year age gap, the younger population had a 5-year gap. An age pyramid was constructed, revealing that the population experienced a decline from ages 41-50, while the younger generations from 0 to 20 remained relatively stable. This observation was further examined in the context of birth and death rates.

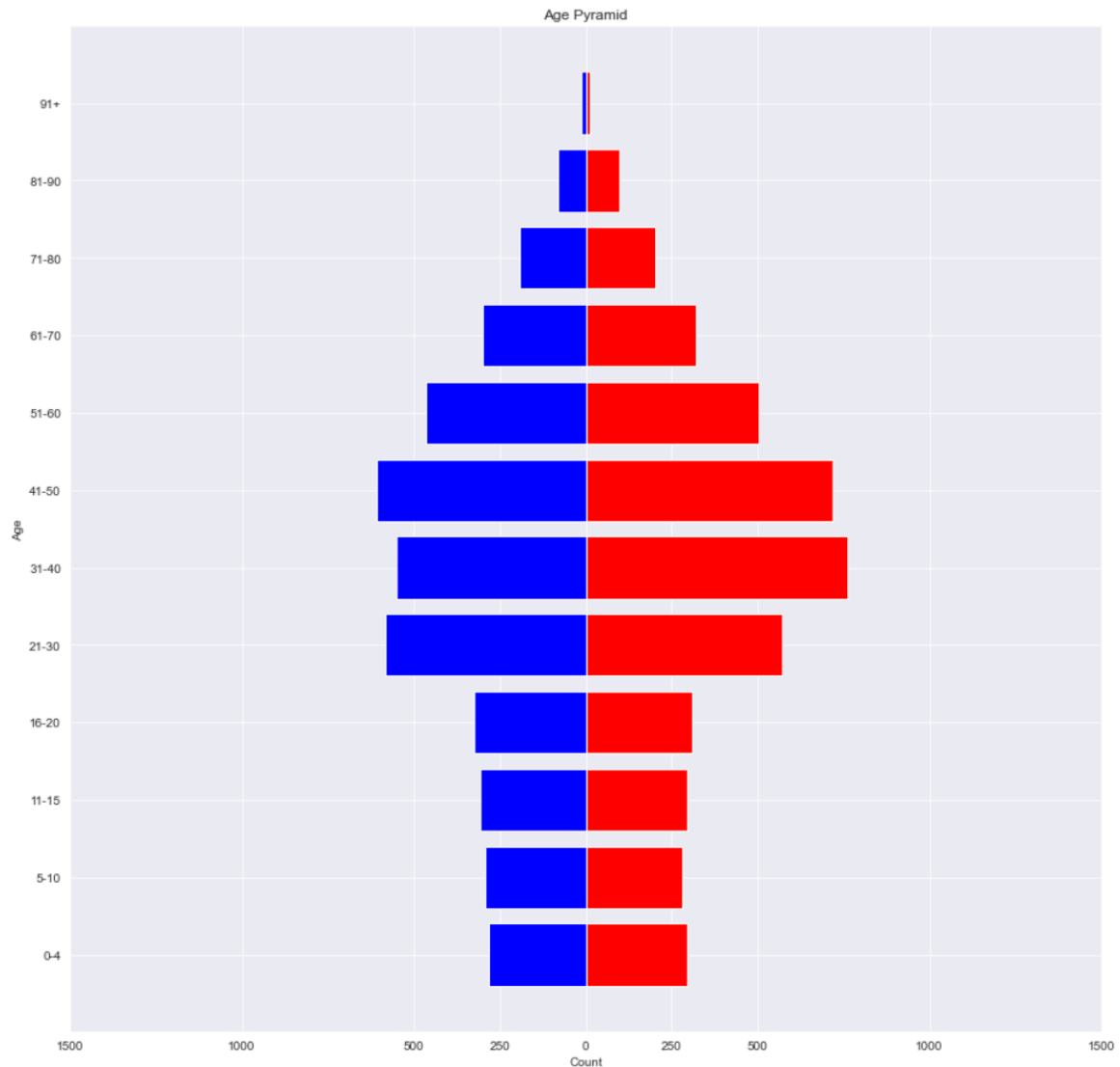


Fig: Age pyramid for the population

Death Rate

Upon examination of the age pyramid, it was observed that there was a reduction in population size in the age range of 61-70. Given that the average life expectancy in the

UK is approximately 80 years (UKHSA, 2021), it is reasonable to infer that this reduction is due to mortality. Additionally, based on the average retirement age of 77, which was calculated, it is plausible to attribute the difference in the Age Group band to mortality, as this particular group is largely composed of long-term residents of the town. The death rate was 198 per 1000 people. This is quite high and will explain the sharp decrease as we move along this age group. It was assumed all other migration causes are negligible.

```
Death rate per 1000
Age Group
61-70    0.620
71-80   -0.225
81-90   -0.215
91+     -0.154
dtype: float64
Death rate
Age Group
61-70    620.0
71-80   -225.0
81-90   -215.0
91+     -154.0
dtype: float64

: # Calculating final death rate per 1000 as average of the last three
death_rate_per_1000 = age_diff_dr.tail(3).mean()
print(f' Death rate per 1000 is {death_rate_per_1000 * -1:.2f}')

Death rate per 1000 is 198.00
```

fig: Death Rate calculation

Birth Rate

With regards to the birth rate, it is defined as the number of births per 1,000 individuals in a specified population within a particular time frame. To calculate the birth rate, the number of births for the most recent three age groups (ages 0 to 2) will be used, and the total population at any given time will be estimated by subtracting the total number of individuals in the preceding smaller age group. In this calculation, it is assumed that there are no other sources of population increase or decrease.

Birth rate two years ago is 11.95 per 1000 people.
Birth rate for one year ago is 10.14 per 1000 people.
Birth rate for current year is 10.62 per 1000 people.

Fig: Birth rate calculation

From the fig above the birth rate across three years are almost the same hence the reason while the population pyramid for the younger population was identical.

EXAMINING THE NUMBER OF UNIVERSITY STUDENT AND RETIREMENT AGE

Upon examination of the population data, it was observed that the number of university students in the town was relatively high at 533 individuals. This finding suggests a high rate of commuting out of the town for educational purposes, as well as for other occupations that require going outside the town for work.

Additionally, it was discovered that the average retirement age of the population was 77 years old. Given the sizeable portion of individuals who are approaching this age, it may be beneficial for the town to make adequate provisions for the needs of its aging population. This may include the creation of social support programs and facilities that cater to the elderly, such as nursing homes or assisted living facilities. Such measures would ensure that the town's retirees are well-catered for and can enjoy a comfortable retirement this can also reduce the death rate.

EXAMINING UNEMPLOYMENT TREND

To provide a more comprehensive analysis, a new column called "employment status" was added to categorize all occupations into Employed, University student, Student, Child,

Retired, and Unemployed. The heatmap analysis of the population employment status showed that the majority of working-class individuals were in the age groups of 31-40 and 41-50. Interestingly, the age groups with the highest unemployment rate were also the same as the age groups with the highest employment rates, indicating that the majority of the workforce falls within these age ranges.

The overall unemployment rate in the town was found to be 7.98%. However, when considering the main workforce age groups of 31-40 and 41-50, the unemployment rates were 10.59% and 11.39%, respectively. These rates are quite high and above the ideal rate of 3-5% (Investopedia 2023). Therefore, there is a need to address the high unemployment rate in these age groups to improve the economic growth and stability of the town

Unemployment rate for each age group is Age Group	
16-20	0.63
21-30	6.50
31-40	10.59
41-50	11.39
51-60	8.50
61-70	8.55
71-80	2.78
81-90	6.67
91+	NaN
dtype: float64%	

Age Group	0-4	473	0	0	105	0	0
	5-10	0	0	0	573	0	0
	11-15	0	0	0	599	0	0
	16-20	0	33	0	307	4	291
	21-30	0	828	0	8	75	242
	31-40	0	1172	0	2	139	0
	41-50	0	1175	0	0	151	0
	51-60	0	883	0	0	82	0
	61-70	0	407	160	0	53	0
	71-80	0	0	384	0	11	0
	81-90	0	0	168	0	12	0
	91+	0	0	24	0	2	0
		Child	Employed	Retired	Student	Unemployed	University Student
		Employment Status					

Fig: Heatmap of Employment status across age group

EXAMINING RELIGION TREND

In order to better understand the religious demographics of the town Religion adoption rate was calculate by getting the percentage of young adults who took their head of house religion. The adoption rate was found to be 78.22%. With Religion adoption rate of 78.22% it's safe to assume children will take their parents religion when they are adults.

A new column called 'prospective and current religion' was created. This column assigns the head of household's religion for children with same surname and maintain the religion of the adults in the population. The resulting data showed that certain religions, such as Pagan, Quaker, and Orthodoxy, are at risk of disappearing as there are no younger members to continue practicing these faiths. With Religion adoption rate of 78.22% it's safe to assume children will take their parents religion when they are adults

Further analysis of the data revealed that Christianity is the dominant religion in the town, with a significant number of followers. As a result, it is likely that a new church will need

to be built to accommodate the growing Christian population, who may currently be commuting outside the town for worship.

	count	mean	std	min	25%	50%	75%	max
Prospective Religion and Current Religion								
Buddist	3.0	29.666667	18.475209	19.0	19.0	19.0	35.00	51.0
Catholic	1217.0	34.430567	21.179884	0.0	16.0	37.0	47.00	102.0
Christian	2438.0	42.016817	22.949198	0.0	22.0	43.0	59.00	116.0
Jewish	52.0	29.980769	19.347686	0.0	13.5	27.0	43.25	70.0
Methodist	657.0	36.936073	22.107460	0.0	18.0	36.0	53.00	100.0
Muslim	115.0	30.747826	16.411295	0.0	20.5	27.0	39.50	94.0
None	3529.0	35.699065	20.052004	0.0	21.0	34.0	49.00	107.0
Orthodoxy	1.0	45.000000	NaN	45.0	45.0	45.0	45.00	45.0
Pagan	1.0	102.000000	NaN	102.0	102.0	102.0	102.00	102.0
Private	1.0	19.000000	NaN	19.0	19.0	19.0	19.00	19.0
Quaker	1.0	57.000000	NaN	57.0	57.0	57.0	57.00	57.0
Sikh	56.0	25.785714	15.711853	0.0	15.0	26.0	31.00	73.0
Undecided	7.0	9.285714	8.712334	0.0	0.0	15.0	16.50	17.0
undecided	299.0	8.789298	4.930295	0.0	4.5	9.0	13.00	17.0

Fig: Statistic summary of Adjusted Religion

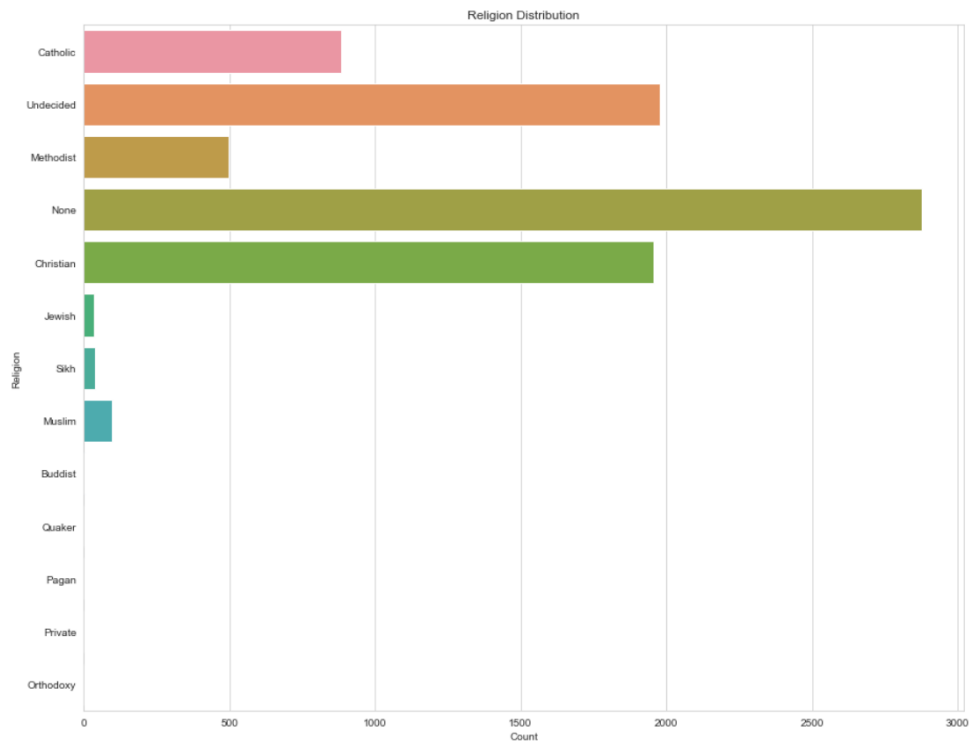


Fig: Bar chart of religion before young population adoption

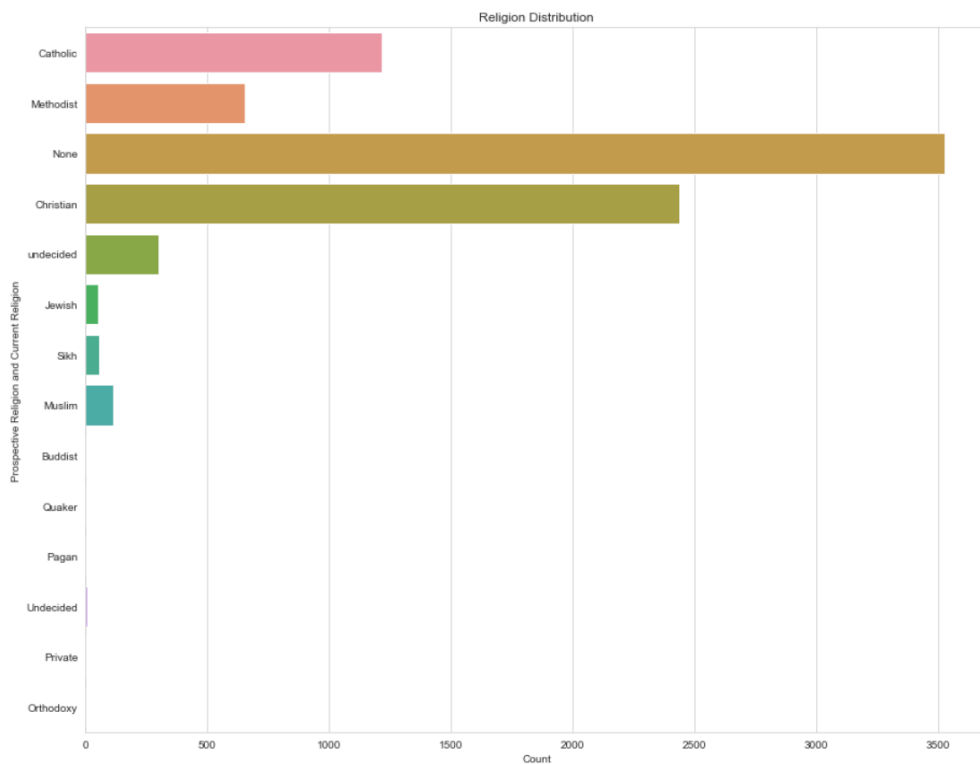


Fig: Bar chart after young population adoption of religion

Examining Occupancy Level

A thorough analysis was conducted to determine the number of individuals per household, and subsequently per street. It was assumed that all streets in the town had the same number of flat and house types. The results indicated that certain streets, such as Davies Avenue and Willow Road, were experiencing significant strain due to a high number of individuals per household and per street. This highlights the urgent need for new housing developments in these areas to alleviate the strain on the current housing infrastructure.

Street	
Davies Avenue	3434.0
Willow Road	2038.0
Harvey Road	1710.0
Cambridge Street	1600.0
ExcaliburBellsby Drive	1303.0
Farthing Crescent	1221.0
Canterbury Well	931.0
Chan Estate	881.0
November Falls	871.0
CherryChestnut Drive	801.0
Singh Burgs	707.0
Aethelred Heights	705.0
Leach Tunnel	680.0
Squirrel Street	639.0
Ward Avenue	571.0
Clark Street	559.0
Pendragon Lane	546.0
Brady Creek	526.0
Murray Drive	515.0
Khan Village	515.0

CONCLUSION AND RECOMMEDATION

From the analysis perform, there is a quite number of people approaching retirement age. Adequate provision should also be made for retirees in the area. Majority of the workforce are in the age group 31-50, which has the highest unemployment rate. The town should prioritize job creation initiatives and provide targeted support for individuals in this age group. Some religions such as Pagan, Quaker, and Orthodoxy will soon fade out due to low adoption rates. The Christian religion has the most members, indicating a need for a new church to accommodate the increasing number of Christians in the town. Some streets are clearly stressed, this indicates a need for new housing development to curb overpopulation. There are also the young adults who might be needing their own housing too.

Recommendations

- (a). High density housing should be built to cater for the areas with high occupancy.
- (b). Employment and training should be invested in to curb the high unemployment in the age group 31-50 which are the main workforce.

Bibliography

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#:~:text=Vulnerable%20children%20across%20England%20and,18%20in%20England%20and%20Wales.>

Wikipedia (no date) *Jediism*

Available online: <https://en.wikipedia.org/wiki/Jediism>

Babynames.com (no date) *Baby girl names*

Available online: <https://babynames.com/girl-names>

UKHSA (2021) *Life expectancy in England in 2020*

Available online: <https://ukhsa.blog.gov.uk/2021/03/31/life-expectancy-in-england-in-2020/>

Gov.uk (no date) *Child Employment: Minimum ages children can work*

Available online: <https://www.gov.uk/child-employment>

Investopedia (2023) *What Is the Unemployment Rate? Rates by State*

Available online:

<https://www.investopedia.com/terms/u/unemploymentrate.asp#:~:text=In%20general%2C%20most%20experts%20deem,on%20what%20constitutes%20healthy%20unemployment.>