

CENSUS PROJECT REPORT

INTRODUCTION

This report contains the findings from the analysis and survey on a moderately sized town with a little over 8000 people living there. The Data is raw, unclean, and to some extent untrue. The overall objective is to analyse the data and make recommendations on the best use of an unused plot of land in the town. The idea in achieving this is to first clean the data, and analyse it by asking questions of the dataset to provide further insights before making any suggestions or recommendations with valid reasons to back up as to why those suggestions were made.

In conclusion, it is discovered, the likely population growth of the town is suggested from a population trend chart which will be discussed later in further detail. There is also an emerging religion which may be making a compelling call to action. Finally, it is a youth-dominated town with the female gender just slightly more than the male counterparts in comparison.

METHODOLOGY

The dataset is from the 1881 Census of the town and given the era we are dealing with, some things will have to be treated in a “simpler” way. For example, the Gender column is restricted to just Males and Females as it was at that time. The data itself is lacking other information or columns that could have helped analyse the population better. There are no details of earning salaries or wages for individuals or the cost of rent or living.

To achieve this project analysis. This data was cleaned and analysed on the Jupyter Notebook framework using Python programming language applying pandas, seaborn and matplotlib libraries.

The complete Jupyter Notebook file is attached for the actual analysis

DATA CLEANING AND INITIAL ANALYSIS

Upon first glance when cleaning the data, it is clear the two stand-out columns requiring attention are the Religion and Marital status columns. These presented a case of over 2000 missing entries for each column. The logical idea was to treat these columns last while going through other columns checking for lies and empty entries as the case may be.

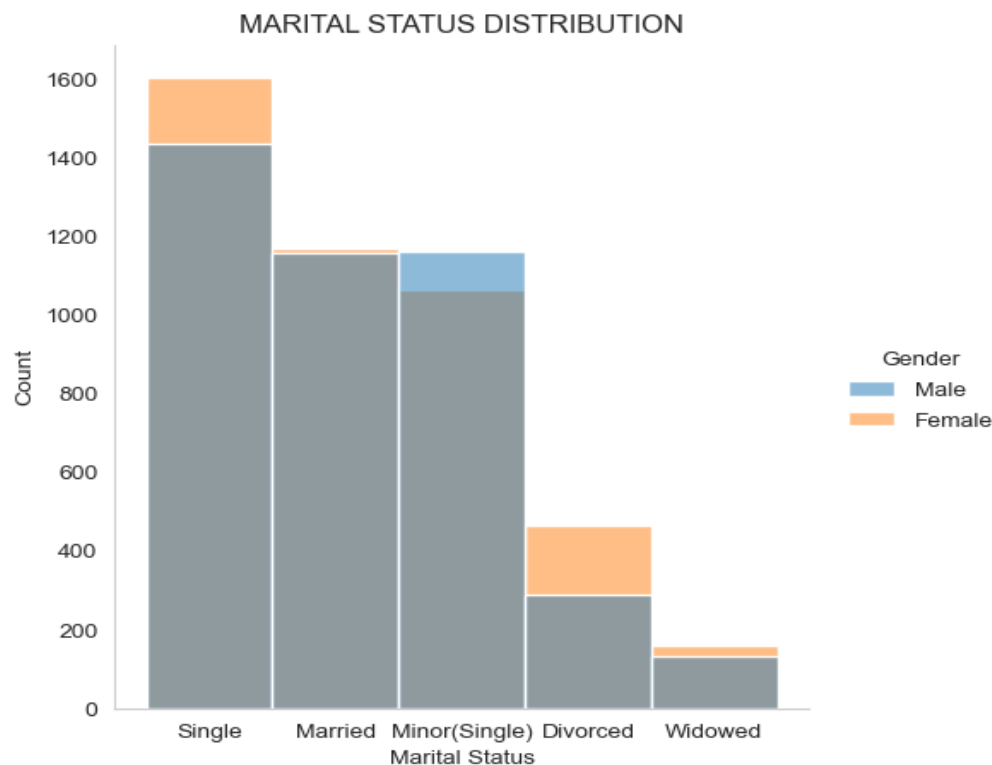
For the Age column, the values were cast to the appropriate data type; Integers. There was a single case of missing string and this was fixed with logic. Given the said person is male, a son and a student. He fit attributed age 16 being the median age for students in the demographic.

In the case of empty first names, they were all replaced with “John Doe” as the three persons were male. A different approach was used for the empty surnames. The ideal solution to handle such missing data is to locate the household of such individuals. In the case that it is a family home, the appropriate surname is given to the individual and in any other case, the mode surname from the population is given to the person. In this singular occurrence, it was “Smith”.

Two cases of missing gender values were treated by locating the individuals and using their first names and relationship to the head of house columns, the genders were fixed

Marital Status

To efficiently handle the missing values in the Marital status column, age has to play a major factor in cleaning. It is important to check for the missing marital values who are under 18 as they are considered minors and not eligible for marriage not at least without parental consent (Marriage Act, 1949). Luckily, the entire population with missing marital values all fall under this category. A further analysis was done to check if they were exceptions, specifically if they were any u18s who may be having a head, husband or wife status as relationship to head of house. They were no such cases. All the individuals in this group were u18, sons, daughters and other values to suggest they still lived with their parents. With this, the group was set to a new marital status value; Minors(Single).



Across all sections, it is clear there are more women represented except for the minors.

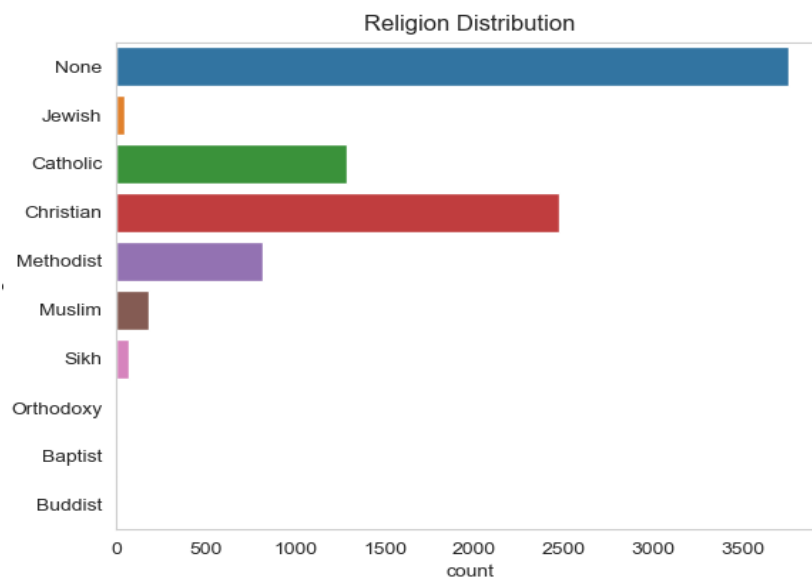
The divorce rate per 1000 of the population is 87. This is in contrast to 270 per 1000 of the population for the Married rate

Religion

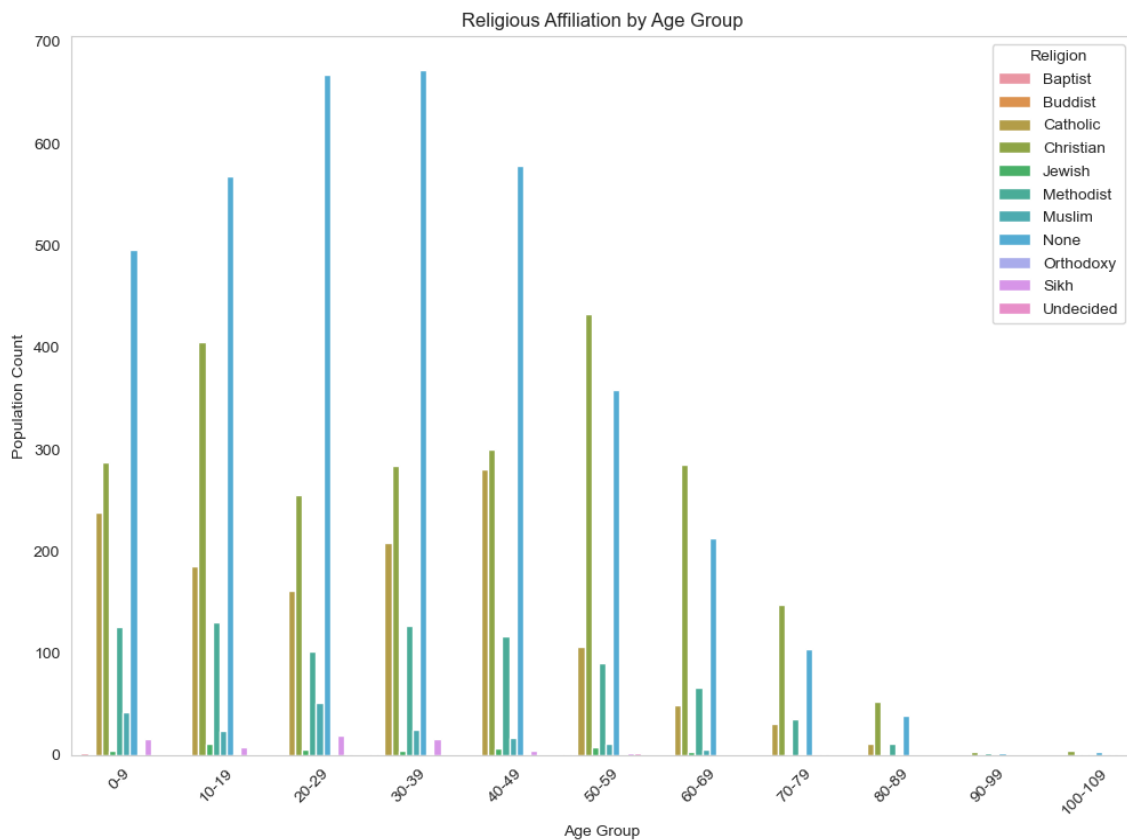
The majority of the Town does not practice any religion. That is in fact, 43% of the entire population. This does not in any way suggest it is a Town with “Non-believers”. The presence of a Catholic church in the town seems fair as the Catholics are the largest religious group in the town. Cleaning the Religion Column with over 2000 missing values is not as straightforward as it may look at first. There are so many factors to consider in handling this situation, more importantly on sensitive information such as Religion. The idea in handling this was to consider the head of home for each household. This is key as it is believed the rest of a house should follow the beliefs or in this context, Religion practised by the head of the house. A function was written to check for individuals who are not Head of the house and assign the religion of the head of the respective houses to them.

This operation sheds more truth on the population as results show, the largest religious followership in the Town is Christians. 29% of the population are Christians, this is followed by the 15% represented by the Catholics.

They were certain individuals who blatantly refused to give a religion simply saying “Nope” and some others “Undecided”. While everyone has a right to his or her own choices. After observations, showing the entire household gave these uncommon answers, the ideal action is to change all those in this group to “None” as there is no possible way of knowing or imposing a belief on them.

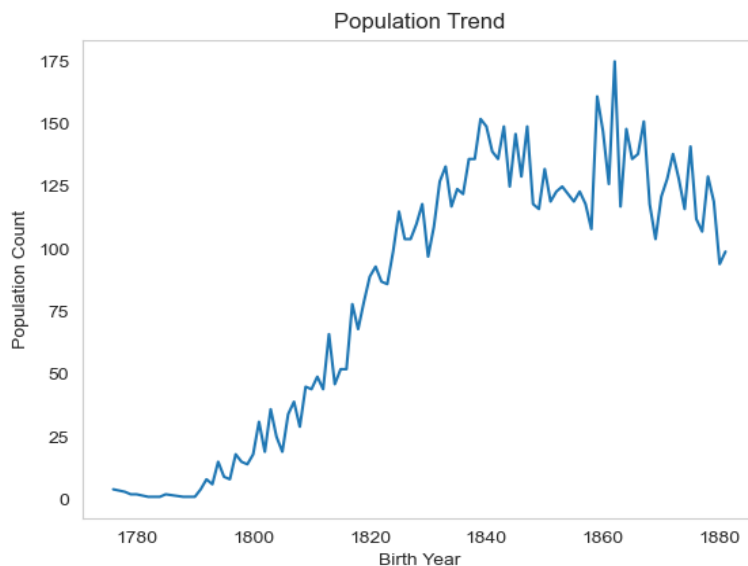


This situation may be calling for a second religious building in the community. This will have to be backed by further analysis. For one, Religious affiliation by age group will be considered and the population growth as well.

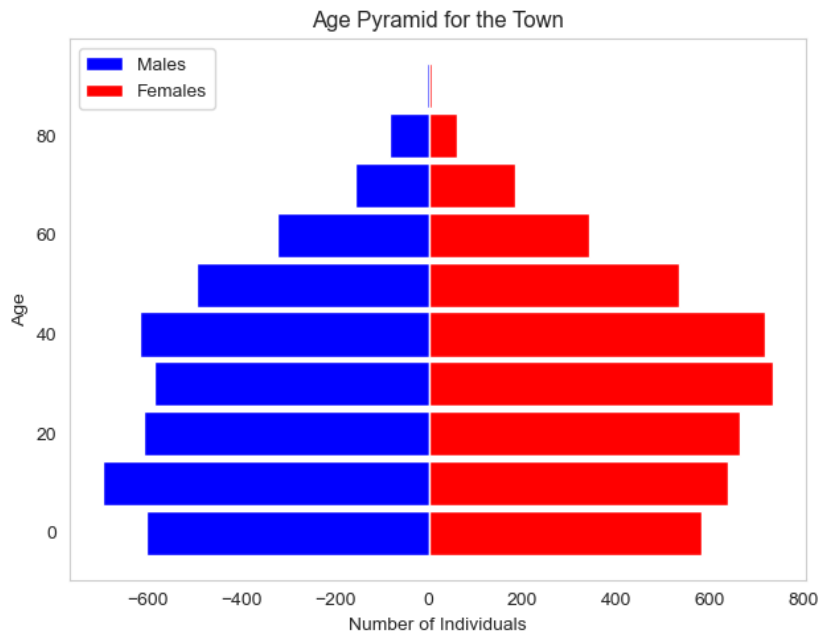


Age Trends

To further understand the population, A population trend chart will be visualized to show us what shape and direction the demographic of the town is taking. To achieve this, a new column will be created to show the Birth Year of each individual. This is achieved by subtracting the current year (1881) from the age of every individual.



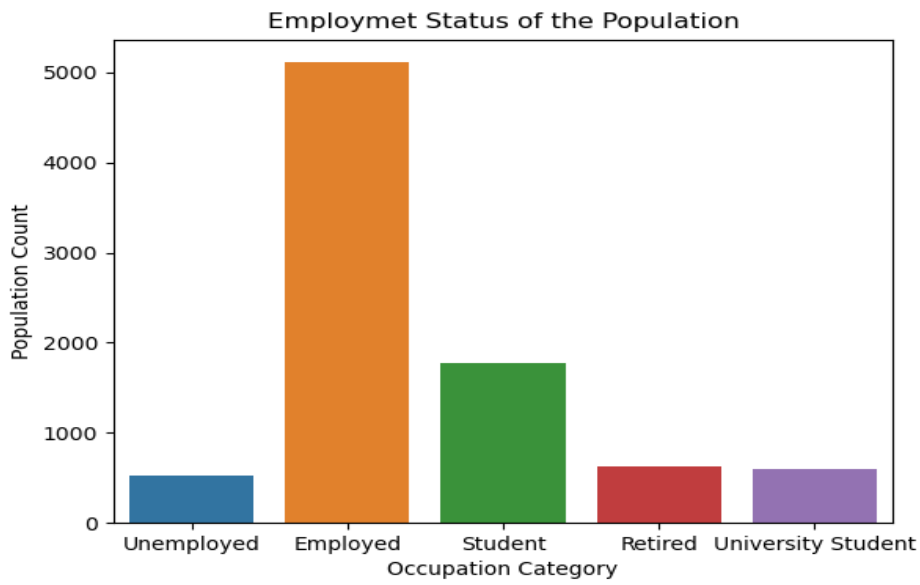
The Population graph above suggests a likely increase or growth in the population of the town going by the trend. This is also backed by the Age Pyramid below. It shows the population ages well. The age chart begins to shrink as it progresses to 60 years and above. The majority live well in their late teens into their late 50s. There also seems to be a balance between the ageing progression of both genders.



Occupation

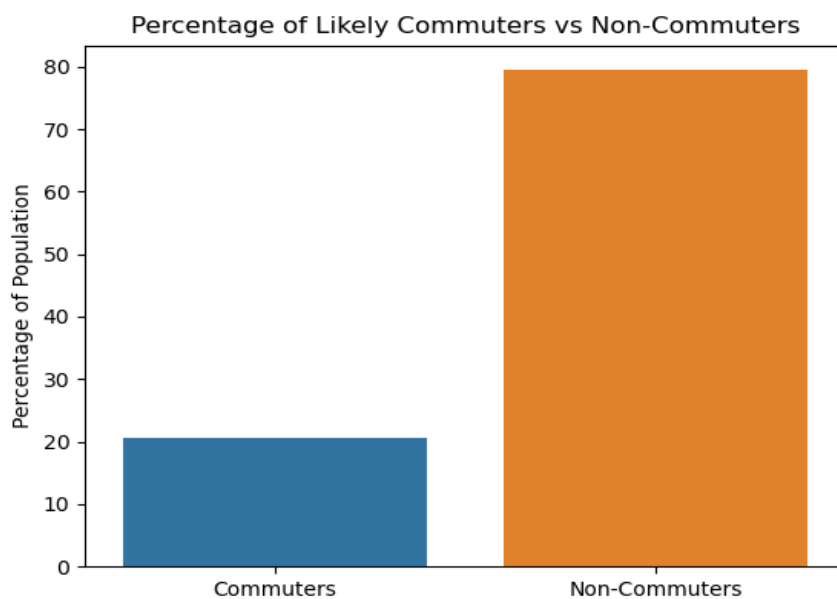
In handling the population column, it is important to correct some notions. The data presents a group of individuals 66 years and above with an unemployed status. Retirement age is 66 According to the UK State Pension (2019). All ages 66 and above with an unemployed status are changed to retired.

Having segmented the Occupation of the Population into sectors. An analysis is carried out on the Employment vs Unemployment status of the town. 6% of the employable population of the town is unemployed This is considered high (Investopedia Team, 2023). Also, 7% of the population is retired.



Commuters

To Identify commuters in the population, the occupation column is thoroughly examined, and Keywords are identified to suggest job roles of people who are likely commuters. One of the possible options for the unoccupied piece of land is to build a train station to ease road movement and also consider the university students who have to travel to the nearby towns. These keywords were meticulously chosen and should suggest regular and likely commuters among the population. This was used to write a function to create another column which contains Boolean values(True or False) for every individual in the Town. The true means, the individual is a likely commuter while the False shows he or she is not.



From the above chart, it can be seen only about 20% of the community are likely commuters. This includes University students and Professors amongst many other job titles in the keyword.

Infirmity

There are minute-to-none infirmities on a larger scale in the town with over 99% of the Town with None status for Infirmity

Housing occupancy Level

Results show the average occupancy level of the population is roughly 3 people per house

INSIGHTS AND RECOMMENDATION

The question beckons, what should be built on the empty plot of land? Secondly, what department should the government invest more in?

From the analysis gathered, 20% of commuters at this current point in time is not enough to justify the construction of a Train station. What this suggests on the other hand is less road traffic and ease of access to travel in and out of the Town via the road network.

What appears to be in demand in the Town is the need for a new religious building. A Pentecostal church probably. This is not entirely needed even though it is backed by almost 30% of the town who declared Christianity as a Religion. Critically thinking, the people who declared Christianity as a Religion may also be Catholic churchgoers. The presence of a Catholic church and not a Pentecostal one may not be totally out of place. The Pentecostal church did not come to the fore until 1901 (Pew Research Centre, 2006)

A case could be made for the Methodist members but there are more pressing issues.

At this point, given the number of Family housing settings in the population data and the housing occupancy level of 3 people per house, suggests the building of more low-density housing should take priority. This is further backed by the population trend showing the future increase in the size of the town. There is a need for more family housing.

Secondly, 7% of retired people in the population with indications suggesting more people in the demographic will be added to this number in the near future suggests that investment in Care homes should also take precedence. A close call would be retraining schemes and tackling the unemployment rate which is roughly 6% of the town but the need to prepare for the inevitable which is caring for the elderly should be slightly prioritized at this point over the employment status of a few which is rather temporary.

Bibliography

State Pension Age UK (2019). [online] Available at: <https://www.ageuk.org.uk/information-advice/money-legal/pensions/state-pension/changes-to-state-pension-age/#:~:text=The%20State%20Pension%20age%20is,again%20from%206%20May%202026>. [Accessed 28/04/2023]

Investopedia (2023) [online] Available at: <https://www.investopedia.com/terms/u/unemploymentrate.asp#:~:text=In%20general%2C%20most%20experts%20deem,on%20what%20constitutes%20healthy%20unemployment>. [Accessed 28/04/2023]

Pew Research Centre (2006) [online] Available at: <https://www.pewresearch.org/religion/2006/04/12/the-new-face-of-global-christianity-the-emergence-of-progressive-pentecostalism/#:~:text=Where%20did%20pentecostalism%20actually%20start,Fox%20Parham's%20Bethel%20Bible%20College>. [Accessed 28/04/2023]

Marriage Act (1949) Section 3 [online] Available at: <https://www.mountfordchambers.com/marriage-a-potential-solution/#:~:text=Changes,relation%20to%20minimum%20age%20requirements>. [Accessed 28/04/2023]

