**CENSUS PROJECT REPORT**

**1.0 Introduction**

This is a census report of an imaginary modest town, located between two cities connected by motorways. The dataset for this project was generated through the Faker package in python. It was designed to model the 1881 census data in the UK.

Since the early 1800s, decision-makers have understood that an accurate census can yield useful information that goes beyond the simple fact of how many people live in each state to enhance the decision-making process. (Council of Economic Advisers, 2000). The aim of this report is to provide necessary statistics, analysis, and recommendations to assist the government to make strategic decisions and develop policies. These policies may include, the kind of infrastructural facilities needed to be developed for the community, any profitable investment that will benefit the town. To achieve these, a thorough cleaning of the raw data (census\_11\_2022) was done.

**2.0 Preparing the Data for Analysis (Cleaning)**

To prepare the data for analysis, it was cleaned to correct any error which includes handling missing data and error entries. These errors can pose a great threat to the accuracy of the analysis and ultimately affect the recommendations to be made. The data was loaded through the pandas library in python for cleaning and analysis (the Jupyter Notebook code is attached). The dataset contains 12 columns and 8485 rows.

**2.1 The “Unnamed: 0” Column**

This column is just a serial number for each of the rows (entries). It is similar to the pandas index and it has no effect or contribution to the analysis. Hence, it was dropped entirely from the dataset.

**2.2 House Number and Street**

These columns have no missing values nor error input. Also, they are in the right data type (object). All looks good; hence, no changes were done to it.

**2.3 Age**

The age column seems to contain the most errors in the dataset. These errors include entry of the age in word instead of number, float entries and an instance of no entry at all. The no entry was handled by inferring from the individual record, which reveals the person is a male, partner to the head of the house (who is a female). From this inference, it is safe to assume this person should be in the same age bracket as the head of house. so, I assumed an age slightly above that of the head of house, 20years. The entries that are in word were replaced with corresponding numbers and the floats were converted to integers, age in years should be a rounded figure. Also, age data type was converted from string to integer.

**2.4 Relationship to Head of House**

The only error spotted here was a particular individual with no entry. Inferring from the household (the house number and street) of this individual shows the individual is a female, with the same surname as the head of the house, both has a marital status of being married and are in the same age bracket (just the two of them are in the household). From this, it is ideal to assume that the individual is the wife of the head of the house.

**2.5 Marital Status and Gender**

The errors in these columns are similar, there are respondents that entered their entries as the first letter of the intended entries, letter “D” as “Divorced” and “M” as “Male”. Rows with these errors were replaced with the intended word. I believe this kind of errors were due to the fact that respondents were given the liberty of filling their answers rather than given the choices to choose from. Also, all individuals under the age of 18years have a marital status of NAN, these entries were left unchanged because these people were considered as children (NSPCC, 2022) . Although these people can be considered as single but single was used for adults old enough to be married but are not, neither divorced nor separated. The legal minimum age to be married in the UK is 18 years (Marriage Act 2022)

**2.6 Infirmity**

The only noticed error with this column is empty entry. Inferring on the household of the people with no infirmity entry shows they are all in the households were all other member has a religion value of “None”, that is, they have no religion. Therefore, the empty entries were replaced with “None”.

**2.7 Religion**

There are individuals with “Nope” and “None” religion values. These two basically means the same thing, therefore, the “Nope” entries were replaced with “None”. Also, all individuals below 18years have religion value of NAN. A child may choose their own religion if they have sufficient understanding (Citizen Advice, 2022). Although, these can easily be set to “None” or assumed the religion of their parent or elders in their household, but it will be difficult to justify such assumption. Nonetheless, people with religion value of NAN and above 18years were changed to “None” because they are matured enough to have or not have a religion. Three religions ( Sikh, Private and Jedi) were changed to none. Jediism is a philosophy, and in some cases tongue-in-cheek joke religion (Wikipedia, 2022)

**2.8 Occupation**

The occupation column has the most unique values. The first error spotted was empty or no entry, this was done using the “.isspace()” method in pandas (just one entry). Inferring on the household of the individual, his siblings (the same surname) in the same age bracket as him were university students, therefore, it was replaced with “University student”. A new column name “Employment Category” was created from the “Occupation” column. This was done to aid analysis in order to determine the unemployment rate of the town. The unique values of this employment category was set to “Child” if the occupation is child, “Student” if the occupation is university student or student, “Retired” if the occupation is retired in any occupation, “Unemployed” if the occupation is unemployed and “Employed” if the person is still working.

Generally, inferring from individual columns and sub-setting on them helps in cleaning the data, especially where we have empty or no entries. There are two individuals who are aged below 18years and are head of the house. According to the census guide, only individuals above 18years can be Head, that is those who filled the census details.

The first person was aged 17years, married to individual of 19years (same surname). This row was left unchanged because it is legal to be married at that age(Marriage Act 2022), and the assumption was that the individual is probably some months away from 18years, hence why she was allowed to fill the survey. The second person was aged 16years, single but has a daughter. This household was dropped as it deemed insignificant to impact the accuracy of the analysis to be done.

There are people who are over 65years, and their occupation was entered as unemployed. The UK pensionable age is around 65years, so their occupations were replaced with Retired.

**2.9 Outlier**

**Chart, box and whisker chart

Description automatically generated**

One of the ways to detect outliers is through visualization. Here, I used box plot to determine if there are any outliers in the age entries. Although the box plot shows some few outliers in the age, the maximum being 105years. I would not consider an age of 105years as being abnormal or error entries, it is possible for someone to attain that age. So, I left it unchanged, concluding there are no outliers.

**3.0 Detailed Analysis**

**Table

Description automatically generated with medium confidence**

The data information after cleaning

**3.1 Age**

**Chart, bar chart

Description automatically generated**

The analysis of the Population pyramid figure above shows that are a greater number of middle-aged people (20-60years) compared to young people (below 20years) and older people (above 60years). Unfortunately, in the future, this will not look good as some of this people will be retired or be in their old age. Generally, we have more female, 4416 than male, 4065 in the population

**3.2 Birthrate and Deathrate**

The crude birthrate decreases in the five years period calculated. This was calculated using the number of children of 0years for present birthrate and children of 4years old for five years prior. This was done to estimate the difference in the rate during this five years period. From the calculation, the birthrate decreases from 12.85 to 9.67 during this period. This indicates a drop of about 3 children per thousand in five years. Hence, the population is not growing even though there are more young people in the town.

Table

Description automatically generated

Similarly, the total deathrate was computed across age groups above 65years. The table above shows the number of deaths per year as well as the deathrate. The total deathrate across these groups was computed approximately to be 15.2 death per thousand household.

**3.3 Commuters**

Commuters are people who live in the town and have high probability of travelling out of the town on a regular basis. Since there are no universities in the town, all university students have been identified as commuters. Also, there are some people with occupation as student who are expected to be in the university in the next 5years (people age 12 years and above) , these people have also been identified as potential commuters. Finally, there are certain employed individuals whose occupation will not allow them to leave the town (e.g. child, primary school teacher or community pharmacist). Nevertheless, we have no information about professions or commercial companies or organizations that are available in the town. since the town is a relatively small town, it is safe to assume half of the employed people work outside the town. From the above assumptions, the total number of commuters was estimated to be 3665 people. This is about 43.2 % of the population.

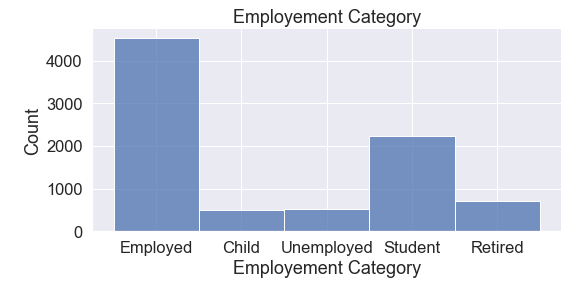
**3.4 Immigrants and Emigrants**

Immigrants and emigrants are very difficult statistics to compute from the dataset we have but an approximate assumption will assist in determining if the town is growing or not. The assumption I made is that, all the immigrants are the lodgers and visitors in the town. Although, the problem with this assumption is that there may be a fraction of the visitors and lodger who will decide not to live permanently in the town but this is difficult to confirm.

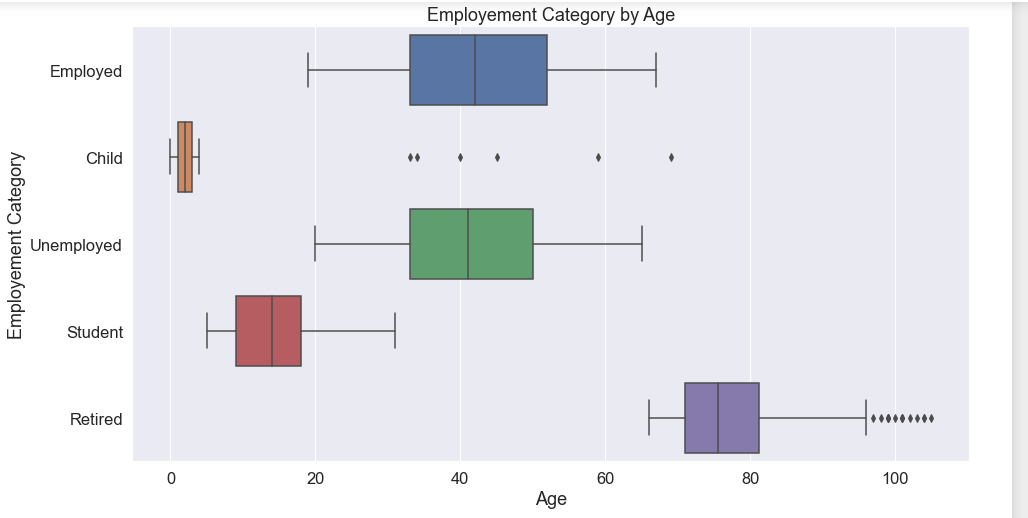
Similarly, emigrants were calculated from difference in male and female divorcees. By these assumptions, the immigration rate is 46.69 per thousand while the emigration rate is 21.9 per thousand household in the town.

From this, a comprehensive understanding of the population growth can be determined from the birthrate, deathrate, immigration rate and emigration rate. This was determined to be 19.2 per thousand household.

**3.5 Employment trend**



From the chart above, it is obvious that the majority of the population are employed while the second highest employment category is students (both university and non-university students).



Majority of the employed and unemployed people are middle class(20-60years) while all the retired are above 60years of age.

**3.6 Religion**

Table

Description automatically generated with medium confidence

Chart, bar chart

Description automatically generated

None religion value account for 33% of the population which indicate majority of the people in the town have no religion. From the box plot, about 50% of the people with religion are between 30 to 60years of age. There are more older people with Christian as religion that any other religions.

**3.7 Marital Status**

**Chart

Description automatically generatedChart, box and whisker chart

Description automatically generated**

From the box plot, virtually all the marital status categories spans from young age till old age. Although there are more widowed old people than younger ones. Similarly, from the scatter plot, there are more female divorcee than male.

**3.8 Housing Demand**

**Table

Description automatically generated**

There are 2819 unique households in the data, the table above shows the average occupancy which is assumed to be the mode of the household in each street. Some households have a lower occupancy that the average occupancy, this is an indication of a low occupancy rate.

**4.0 Recommendations**

Obviously, the town has a good employment rate based on the number of employed people being highest in the town, but it is not known if majority of them are working in the town. What is known is that majority are of middle class age. In this case, I will recommend more employment opportunities in the town and trainings for unemployed people. This will also benefit students in the community and encourage them to stay after their studies.

Due to the fact that majority of the people in the population are middle aged, in the future, some of this people will be in there old age adding to the existing old people. Hence, I will encourage allocating more funding to old age care and end of life support.

Commuters are about 43% of the population, this is on the high side. Since the town is only connected by motorways to the neighboring cities, I will suggest building an alternative transport system, e.g. a train station. This will not only decrease the travel time in and out of the town but it will also reduce the pressure on the motor ways, thereby reducing road traffics as well as the maintenance. Since the occupancy rate of the town is low, it shows some house are under used. So, I wouldn’t demand the construction of new housing in the town.

Although the dataset indicated about 33% of the population has no religion, Christianity accounts for around 23%. While catholic being the third on the list already has a place of worship, the only religious development I will recommend is the building of a church. Other religions are less than 10% each.

I anticipate a growth in the number of old people 10years in the future, this demands for a new healthcare building if there is none or expansion of the existing one to accommodate the surge.

Majority of the population are between 20 to 50 years; this indicates the town has a lot of people within the working age. A profitable investment I will recommend is manufacturing, they have the workforce. Manufacturing will also attract investors.

**5.0 Reference**

Citizens Advice (2022). Young people's rights.

Available online: <https://www.citizensadvice.org.uk/law-and-courts/civil-rights/young-people-s-rights/>

[Assessed 10/11/2022]

Council of Economic Advisers (2000, April 1) The Uses of Census Data: An Analytical Review.

Available online:

[https://clintonwhitehouse5.archives.gov/WH/EOP/CEA/html/censusreview.html](https://clintonwhitehouse5.archives.gov/WH/EOP/CEA/html/censusreview.html  )

[Assessed 10/11/2022]

Marriage And Civil Partnership (Minimum Age) Act (2022), Division 2.

Available online: <https://www.legislation.gov.uk/ukpga/2022/28/notes/division/2/index.htm> [Assessed 21/11/2022]

National Society for the Prevention of Cruelty to Children, NSPCC (2022) Children and the Law.

Available online: [https://learning.nspcc.org.uk/child-protection-system/children-the-law#skip-to-content](https://learning.nspcc.org.uk/child-protection-system/children-the-law%23skip-to-content%20)

[Assessed 21/11/2022]

Wikipedia. (2022.). Jedi census phenomenon.

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

[Assessed 26/11/2022]