

## **1.0 INTRODUCTION**

The project provides a mock census of an imaginary modest town, with the objective of carrying out a statistical and classical analysis of the given data and then provide recommendation for growth. Having provided with a plot of land and funds; good decisions to be made on how the land will be maximized for infrastructural development and funds wisely allocated.

The rubric for this decision is not limited to:

- Unemployment trend
- Birth and Death rate
- Commuter level in the town
- Population trend: Constrictive, Expansive or Stable
- Projected Aged people in the future
- Highest followers' rate for a religion

This report will provide a detailed Exploratory Analysis of the Data; Cleaning of the Data; how software codes and visuals generated from Data will bring prosperity to the town amongst many.

## **2.0 METHODOLOGY**

In the Methodology section is approach followed for the Data exploratory to getting contextual information and then cleaning for Analysis. Data cleaning techniques were explored having imported data into Pandas and insights gotten for further decision.

### **2.1 EXPLORATORY DATA ANALYSIS**

With the Jupiter Notebook, the Census data was imported. The data attributes from respondents includes: First Name, Surname, Street, Marital status, Gender, Occupation, Infirmary, Religion, Age, Relationship to Head of House and House number. These Data attributes have distinct elements attributed to it. The shape of the Census Data was determined with the shape method to get the total number of populate which equals 11,296people. Also, an inappropriate column: 'Unnamed:0' ( extra column for index) was deleted as the Pandas will create its own default index for the Data frame.

In addition, the info method showed the Data types of each column which guide on the inappropriate Data type and the required actions for error handling. Check was also done for the duplicacy.

## 2.2 DATA CLEANING

The importance of properly cleaning data can't be overstated. It's like creating a foundation for a building: do it right and you can build something strong and long-lasting. Do it wrong and your building will soon collapse (Career foundry, 2022). The say from Career foundry has laid the basis on how proper Data cleaning can influence the quality of decision one makes from a Data. In a bid to make a good decision for the town, each Data column was properly handled; all errors corrected. Below table show error discovered with techniques to handle while detailed analysis of the cleaning can be found in the Jupyter Notebook.

*Table 2.1 Data Cleaning techniques on the Data columns*

DATA COLUMN	ERRORS IN COLUMN	TECHNIQUES TO CLEAN DATA
Age	<ul style="list-style-type: none"><li>• Elements are floats and Object type</li></ul>	<ul style="list-style-type: none"><li>• Age column as object types and float assigned the right integer data type</li></ul>
Gender	<ul style="list-style-type: none"><li>• Inconsistent Data and Blank Entry</li></ul>	<ul style="list-style-type: none"><li>• Inconsistent Data made consistent</li><li>• blank entry from three (3) respondents replaced with the Male category upon inspection and judging from the First name and Relationship to head</li></ul>
Marital status	<ul style="list-style-type: none"><li>• Missing Values and Inconsistent values</li></ul>	<ul style="list-style-type: none"><li>• All missing values found in the category of Children and considering being Minors, it was replaced with N/A.</li><li>• Inconsistent entries made consistent</li></ul>
Religion	<ul style="list-style-type: none"><li>• Blank Entry and Missing values</li></ul>	<ul style="list-style-type: none"><li>• Missing value replaced with None Category as appropriate</li><li>• Nan in the Minor Category replaced with N/A</li></ul>
House Number	<ul style="list-style-type: none"><li>• Inconsistent values</li><li>• Data type being Object</li></ul>	<ul style="list-style-type: none"><li>• Inconsistent made consistent</li></ul>

		<ul style="list-style-type: none"> <li>Integer Data type assigned to column</li> </ul>
First Name	No error	<ul style="list-style-type: none"> <li>NA</li> </ul>
Relationship to head of the House	<ul style="list-style-type: none"> <li>Mis-spelt name: Neice</li> <li>Missing Values and case of Child being Head</li> </ul>	<ul style="list-style-type: none"> <li>Misspelt Name corrected</li> <li>Child as head made daughter upon proper check with relationship with similar surname</li> </ul>
Surname	<ul style="list-style-type: none"> <li>Blank entry</li> </ul>	<ul style="list-style-type: none"> <li>Blank entry trace to common surname: 'Stephenson' and then corrected</li> </ul>
Street	<ul style="list-style-type: none"> <li>No error</li> </ul>	
Infirmity	<ul style="list-style-type: none"> <li>Blank entry</li> </ul>	<ul style="list-style-type: none"> <li>Mode was used to fix this; hence, corrected by assigning a None category</li> </ul>
Occupation	<ul style="list-style-type: none"> <li>Wrong and blank entry</li> </ul>	<ul style="list-style-type: none"> <li>Blank entry as underage replaced with Student. Also, for simplicity; a new column: 'Employment category' was created to further categorize the occupation into classes</li> </ul>

With the Data cleaning techniques used on each column, the info method is being called up to give a detailed summary of our Dataframe. Below picture shows us information about the Dataframe which confirms each Data column with right Data types.

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 11296 entries, 0 to 11295
Data columns (total 13 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   House Number                          11296 non-null  int32
1   Street                                11296 non-null  object
2   First Name                            11296 non-null  object
3   Surname                               11296 non-null  object
4   Age                                    11296 non-null  int32
5   Relationship to Head of House         11296 non-null  object
6   Marital Status                        11296 non-null  object
7   Gender                                11296 non-null  object
8   Occupation                            11296 non-null  object
9   Infirmary                             11296 non-null  object
10  Religion                              11296 non-null  object
11  Employment category                   11296 non-null  object
12  Age_group                             11157 non-null  category
dtypes: category(1), int32(2), object(10)
memory usage: 982.1+ KB
```

Fig 2.1 Detailed summary of the Census Dataframe

## 2.3 PACKAGE

The Jupyter notebook an open-source web application which contains inbuilt packages like Seaborn, Pandas, NumPy for manipulation, visualization etc. It was used for this analysis

### 3.0 RESULTS & DECISIONS

Below is summary of results:

#### 3.1 POPULATION DEMOGRAPHY

The Town population makes up 11,926 people comprising of both Male and female in the percentage 48% and 52% respectively. The Mean Age is 35 and Maximum Age is 110 years which is an indication of high life expectancy. The population pyramid below revealed a large proportion of the population is made up of the middle age and children; also, we have almost equal proportion of the middle and the children age group while we noticed a plum in the birth rate between Ages (0 – 8 ) which is an indication of a low birth rate for the year. I can say that the graph is constrictive having a plum in the birth rate and near stationary as we have almost equal proportion of the children and the adult. From the population composition in Fig 3.2, we can also see that the percentage of marriage (25%) is low as compared to singles (35%) which is an indication of lone occupancy within the town.

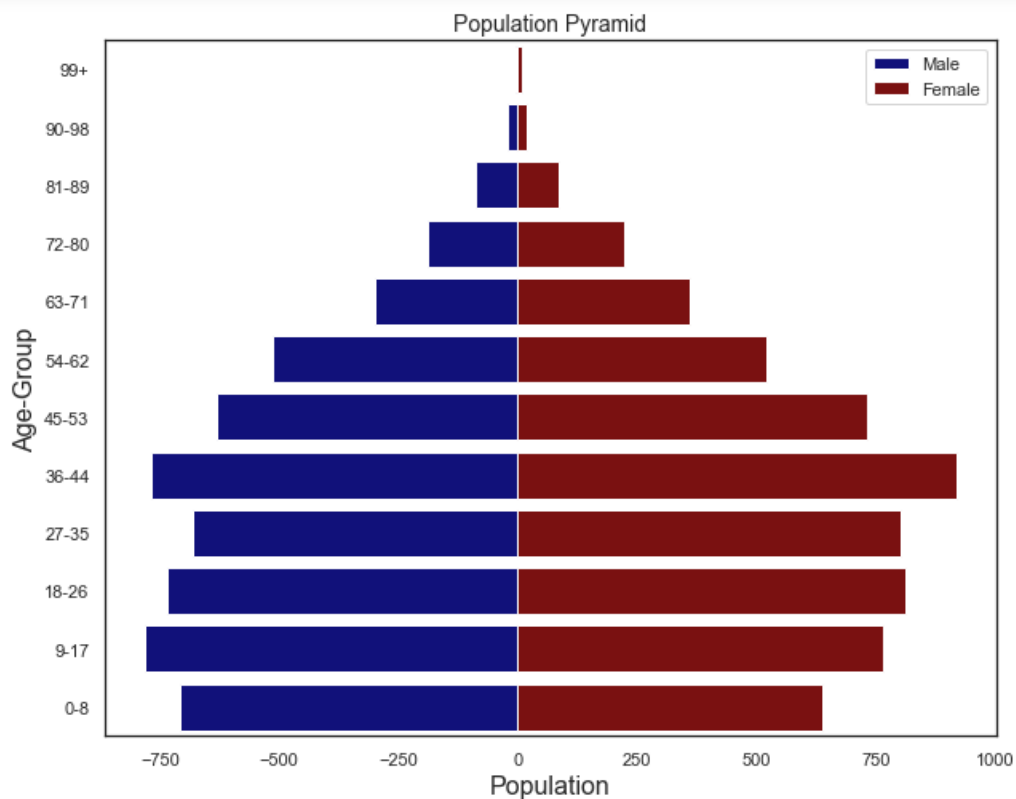


Fig 3.1 Population Pyramid

```

Town population: 11296
-----
The maximum Age: 110
The mean Age: 34.876947592067985
-----
The Female population: 5875
The %Female population: 52.00956090651558
The Male Population: 5421
The %Male Population: 47.99043909348442
-----
The %divorced population: 9.791076487252125
The %Married population:25.53116147308782
The %Single Population: 35.4284702549575
The %widowed population: 3.6827195467422094
-----
The children population: 2889
the Adult Population: 8407
The %Children Population: 25.57542492917847
The %Adult Population: 74.42457507082153
-----
The percentage makeup of the Age group (18 - 25): 12.376062322946176
The percentage makeup of the Age group (26 - 45): 30.771954674220964
The percentage makeup of the Age group (46 - 65): 22.105169971671387
The percentage makeup of the Aged group: 9.171388101983002
-----
Married Female :1447
Married Male :1437
-----
Divorecd Male :452
Divorecd Female :654
-----
Female single :2113
Male single :1889
-----
Male Widowed :154
Female Widowed :262

```

Fig 3.2 Population Composition

## 3.2 UNEMPLOYMENT TREND

The headline measure of unemployment for the UK is the unemployment rate for those aged 16 and over. Unemployment rates are calculated, in accordance with international guidelines, as the number of unemployed people divided by the economically active population (those in employment class plus those who are unemployed class) (ONS,2022).

From Employment/Unemployment statistical table gotten from our numerical analysis(code), as seen is the unemployment rate hitting 11%. Based on ONS analytics on the Unemployment rate in the third quarter of 2022, I would conclude this is a high level of Unemployment as it is almost thrice the estimated figure.

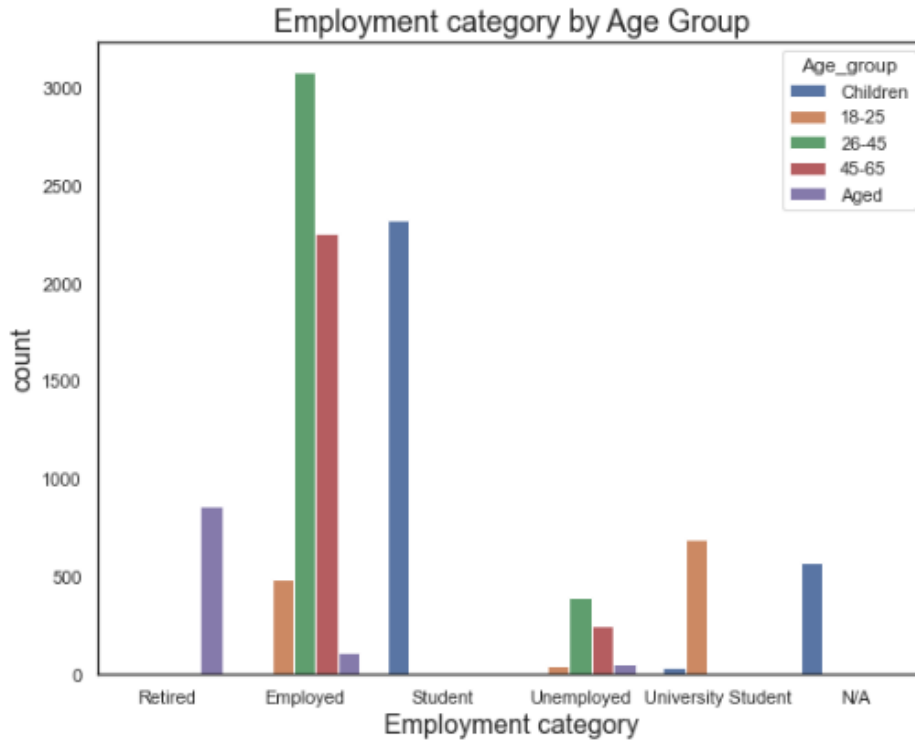


Fig 3.3 Employment Category by Age group

Table 3.1 Employment statistical Table

EMPLOYMENT CATEGORY	CATEGORY PERCENTAGE OF TOTAL POPULATION	EMPLOYEMENT RATE
Retired	7.6%	NA
Children	6.3%	NA
University Student	6.4%	NA
Unemployed	6.5%	11%
Student	20.5%	NA
Employed	52.0 %	NA

**DECISION:** On what should be invested on, funds should be allocated into employment and training to drastically reduce the unemployment rate; hence, bring prosperity to the town.

### 3.3 RELIGION AFFILIATION

The statistical description of the Age across religion and the boxplot below gives a detailed analysis on the rank of followers for the various religion. As can see, the town has many followers in the Christian and catholic category, 2426 and 1158 people respectively; however, a true indication to judge if there tend to be future followers of this religion is by the median Age followers.

The median Age follower for Christian and catholic is 49 and 41 (Middle age) and the Maximum age follower for both is 104 and 104 respectively. We can conclude that Old and Middle age practice this top ranked followed religion; hence, a phase out in the future as the younger generation won't show interest in the practice

	count	mean	std	min	50%	max
Religion						
Agnostic	1.0	42.000000	NaN	42.0	42.0	42.0
Bahai	1.0	39.000000	NaN	39.0	39.0	39.0
Catholic	1158.0	41.693437	15.010920	18.0	41.0	104.0
Christian	2426.0	48.690025	18.112124	18.0	49.0	104.0
Jedi	4.0	51.750000	15.456929	30.0	55.5	66.0
Jewish	29.0	42.689655	19.411159	18.0	42.0	80.0
Methodist	768.0	44.985677	16.751456	18.0	42.5	98.0
Muslim	142.0	38.492958	16.760620	18.0	32.0	91.0
N/A	2947.0	9.083135	5.232368	0.0	9.0	24.0
None	3747.0	41.862557	16.681753	18.0	40.0	110.0
Orthodoxy	1.0	29.000000	NaN	29.0	29.0	29.0
Sikh	72.0	32.916667	10.867642	19.0	30.5	71.0

Fig 3.4 Religion Data Column Statistical Description



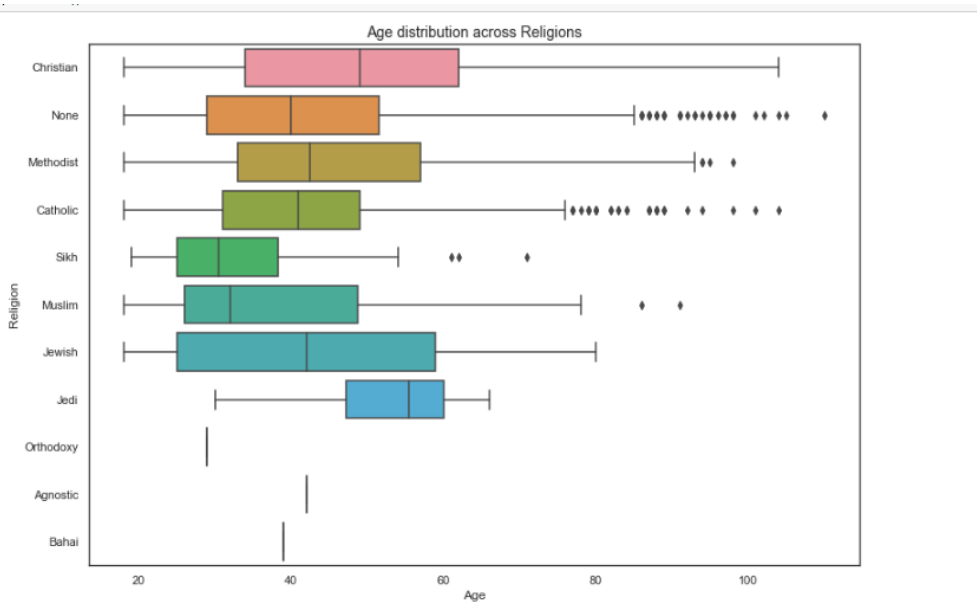


Fig 3.5 Boxplot showing Age distribution across Religion

**DECISION:** what should be built on an unoccupied plot of land?

From the result analysis, religion building isn't an option.

### 3.4 SCHOOL CHILDREN GROWTH RATE

Based on my computation, the population category put into analysis is the Children class. The growth rate between the Present year (Age 0-16) and past One year (Age 1 – 17) was calculated which revealed a negative growth with a rate of -1.38 and a reduction of 38people for each year. For the Census year, the Children population sits at 2674children while projected second, third, fourth- and tenth-year projection are 2637, 2601, 2530 and 2359 respectively which indicates a plum over the year in the number of school children. However, comparing the present year and tenth year projected value, we would notice about 11% value reduction although a plum but not too substantial for that long year.

School Children Yearly Growth	
year1_Projection	2674.525091
year2_Projection	2637.568017
year3_Projection	2601.121623
year4_Projection	2529.732743
year10_Projection	2359.715248

Fig 3.6 School Children Projected Yearly growth

**DECISION:** Based on these facts, I would advise the school funding shouldn't take precedence

### 3.5 EMERGENCY MEDICAL BUILDING

For this analysis, I have factored the vulnerabilities: Crude birth from pregnancies, the children especially within the Age (0 – 5) and the aged class (65 – 110). The percentage growth change from one period to another is calculated using the Uoregon approach:  $\text{Percentage change} = ((\text{Present Value} - \text{Past Value}) / \text{Past Value}) \times 100$  (Uoregon, 2022).

Upon computation, it was discovered that we had a positive vulnerability rate for the crude birth and Child (12% yearly rate) meaning there tend to be more vulnerability by the year; for the aged class is a negative rate as this might reduce due to natural death. Overall, 16% of the total population will cause a shift for emergency building within the town.

**DECISION:** Based on this result, Emergency building will need to be invested on

### 3.6 OLD AGE CARE

From computation in the death rate section in the Jupyter notebook, we calculated a death of 134 people within 1 year with a positive death rate of about 12%. Projected death in the next five year with a constant death rate will amount to 670 people which infers a quick look into investment for the End of life.

Secondly, viewing the population pyramid, we have a larger number of middle class than other class in the population which is an indication that we will have an increased number of aged people in the near future. Conclusively, based on these analysis, end of life care need to be invested in.

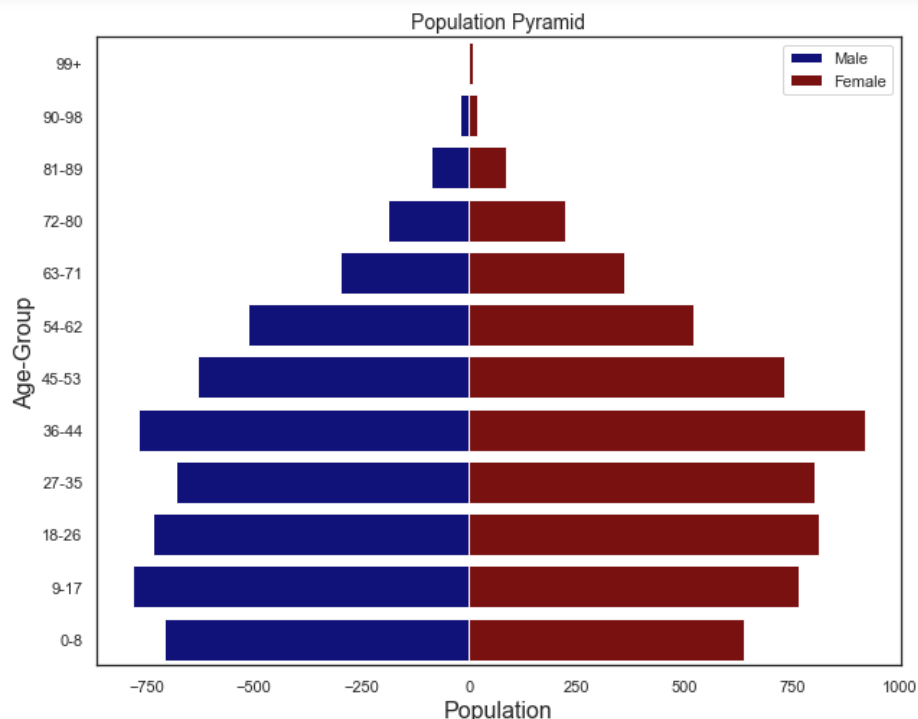


Fig 3.7 Population Pyramid

**DECISION:** Allocation of funds need be channeled into End-of-Life Care

### 3.7 COMMUTERS

Commuters are those who travel between home and work on a regular basis. After a close look into the different employment category list, I discovered about 80% commute to the city which can be approximately categorized as commuters: hence, the University student and the employed fall into the commuter class. From the Jupyter notebook computation, percentage of commuters is calculated by dividing the sum of the Employed and the University students by the economically active people. By computation, the percentage commuters is 85% of economically active class and about 59% of total population

**DECISION:** Transportational mix (Train station) need be given consideration as it is justifiable having about 85% of the economically active people commuting to work

### 3.8 OCCUPANCY RATE

From Computation, a new data column was created which holds the number of occupancies per person as household. Taking the median as the reference, we can see that there are 754household above 4numbers occupancy, leaving us with 2747household below 4numbers occupancy which is about 22% of total household in the town.

**DECISION:** A low density housing is recommended in this case

In summary,

On the Unoccupied Plot of Land, below are recommendation:

- Low Density Houses
- Emergency building
- Train station

What should be invested on:

- End of Life care
- Employment and Training

## BIBLIOGRAPHY

Careerfoundry, 2022.careerfoundry.com. [Online]

Available at:

<https://careerfoundry.com/en/blog/data-analytics/what-is-data-cleaning/#why-is-data-cleaning-important> [Accessed 4 December 2022]

Futurelearn, 2022.futurelearn.com. [Online]

Available at:

<https://www.futurelearn.com/info/courses/introduction-to-environmental-science/0/steps/270677#:~:text=To%20calculate%20the%20Population%20Growth,population%20and%20multiply%20by%20100> [Accessed 8 December 2022]

ONS, 2022.ons.gov.uk [Online]

Available at:

<https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/methodologies/aguidetolabourmarketstatistics> [Accessed 6 December 2022]

Uoregon,2022.Population growth sample. [Online]

Available at:

<https://pages.uoregon.edu/rgp/PPPM613/class8a.htm>

Study, 2022.study.com. [Online]

Available at:

<https://study.com/learn/lesson/fertility-rate-overview-calculation.html> [Accessed 5 December 2022].