

Introduction

This is a data analysis report of a mock census of an imaginary town in the United Kingdom with a population of 8577 in 1881. Towns with population between 5000 and 100000 are considered mid-size towns (Vij et al, 2022), and our imaginary town fits this description. The town has a total of 2991 households in 104 streets with an average occupancy of 3 per households. Persons with ages between 30 -39 (16.37%) and 40 - 49 (16.43%) constitute most of the population at 32.8%. While ages 0 – 9 make up 11.45% of the population, the older population, from 70 and above constitute 7.45% of the population. Economically, while 54.49% of the population are gainfully employed, the unemployment rate is 9.3%. The total student population 25.77 of the town's population and university students account for 23.98% of the student population. The town is sandwiched between two much larger cities and it is connected by motorways to these cities.

The aim of the analysis is to make recommendations for project selection for developments on an unused plot of land. This analysis takes into cognizance of the fact that the council will need to maximize the benefits (social, economic and environmental) from every taxpayers pound (Local Government Association, 2013) and prioritize projects which cost benefits analysis (CBA) is higher (Marcelo et al., 2016; emeritus.org, 2023) keeping in mind that the main source of funding for local government in England are council tax (MHCLG, 2021).

I first explored the dataset, and then proceeded to clean the data. Summary of the data cleaning is provided in the first part of this report and the detailed analysis followed and then final section is the recommendation.

Data Cleaning:

The mock census data was checked for data errors and a logbook of all cleaning undertaken can be found in the corresponding Jupyter Notebook.

1. I used `df[Column_name].unique()` to check for veracity of records in the columns and possible outliers and the checked returned negative
2. I also checked each column for the presence of empty strings exist.
3. Out of the total of 11 columns, House Number, Street, First Name, Gender, Occupation had no issues requiring cleaning.
4. The datatype of the Age column was changed from float to int.
5. Empty strings in the Relationship to Head of column were cleaned by looking at other members living in the same household to make inference.
6. Empty string in surname was fixed by referring to person living in same household.
7. Null rows and empty strings in Infirmary column were cleaned by setting these to 'No_Declared_Infirmary' while null values were handled with `df['Infirmary'].fillna('No_Declared_Infirmary')`.
8. Religion Column: Empty string here was cleaned by setting to 'Unknown' while null values were handled with `df['Religion'].fillna('Unknown')`. Sith was also changed to 'Unknown' while Bahai with one row was left as this religion is known to exist (Wikipedia, 2024).
9. Marital status has 2085 null value corresponding to the total number of rows in the Age column of persons who are less than 18 years. The **Age of Marriage Act 1929**

increased the Age of marriage to 16. **Marriage and Civil Partnership (Minimum Age) Act 2022** made the legal age of marriage to be 18 (Wikipedia, 2024; Legislation.gov.uk, 2022). It was not unlawful to marry below 16 in the UK as at 1881. However, I checked and saw that no one who was under 18 was living in a house with a partner or spouse. I proceeded to update the marital status of all under 18 to 'Single'.

10. Occupation column has values like 'Sub', 'Press Sub', 'Make', 'Copy' and 'Land' which I cannot precisely relate to the actual job and was classified as 'Others' in salary computation. In analysis more columns were, and this column provided basis for some of these new columns. Salary values were derived from Office for National Statistics (www.ons.gov.uk, 2023). Salary value for 'Others' was arrived by calculating a mean of the salary of available salary information. Rows with value as unemployed in this column whose have above 65 in 'Age' column were considered as 'Retired' in all computation.
11. After cleaning, we need additional columns to help our analysis and we created 'Occupancy', 'Birth Year', 'Work status', 'Occupation category' and 'Annual salary'.

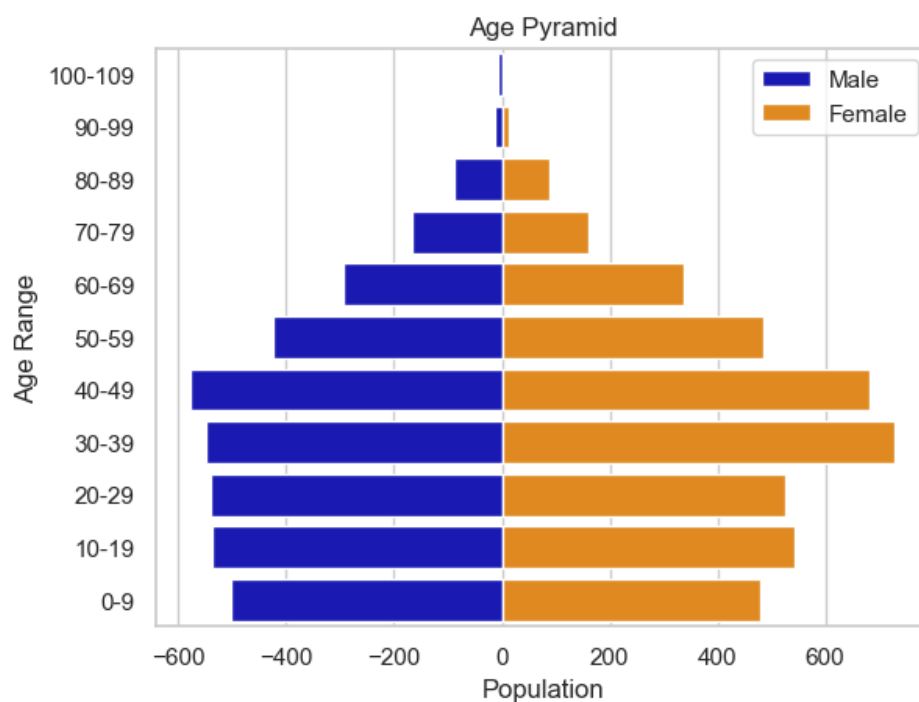
After the cleaning exercise, our data is summarized as below:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8577 entries, 0 to 8576
Data columns (total 17 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   House Number                        8577 non-null   int64
1   Street                             8577 non-null   object
2   First Name                         8577 non-null   object
3   Surname                           8577 non-null   object
4   Age                               8577 non-null   int32
5   Relationship to Head of House      8577 non-null   object
6   Marital Status                    8577 non-null   object
7   Gender                           8577 non-null   object
8   Occupation                        8577 non-null   object
9   Infirmary                         8577 non-null   object
10  Religion                          6445 non-null   object
11  Birth Year                        8577 non-null   int32
12  Age Range                        8577 non-null   category
13  Work status                      8577 non-null   object
14  Occupancy                       2991 non-null   Int64
15  Occupation category              8577 non-null   object
16  Annual Salary                   2895 non-null   float64
dtypes: Int64(1), category(1), float64(1), int32(2), int64(1), object(11)
memory usage: 1022.4+ KB
```

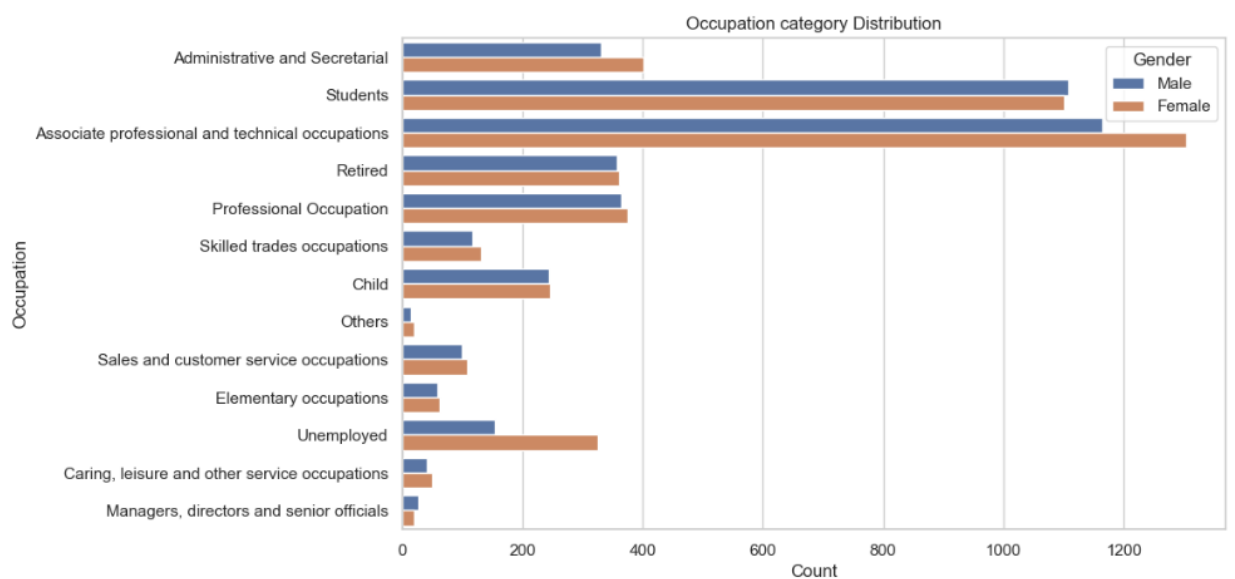
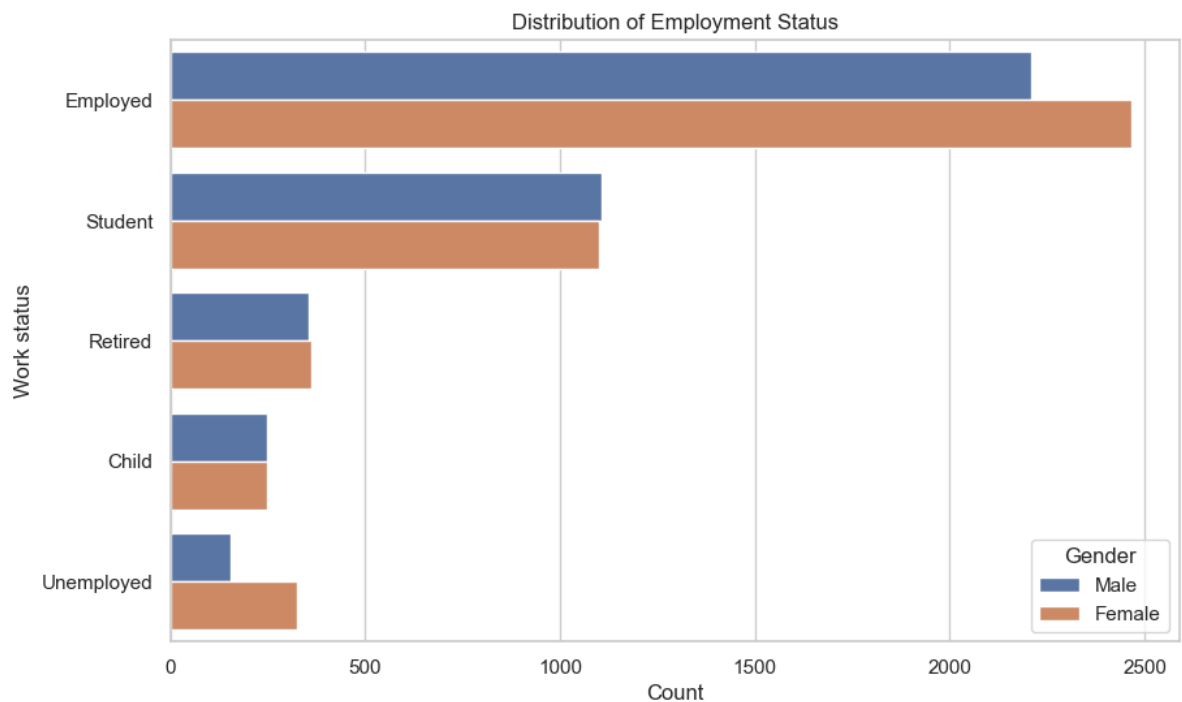
Data Analysis

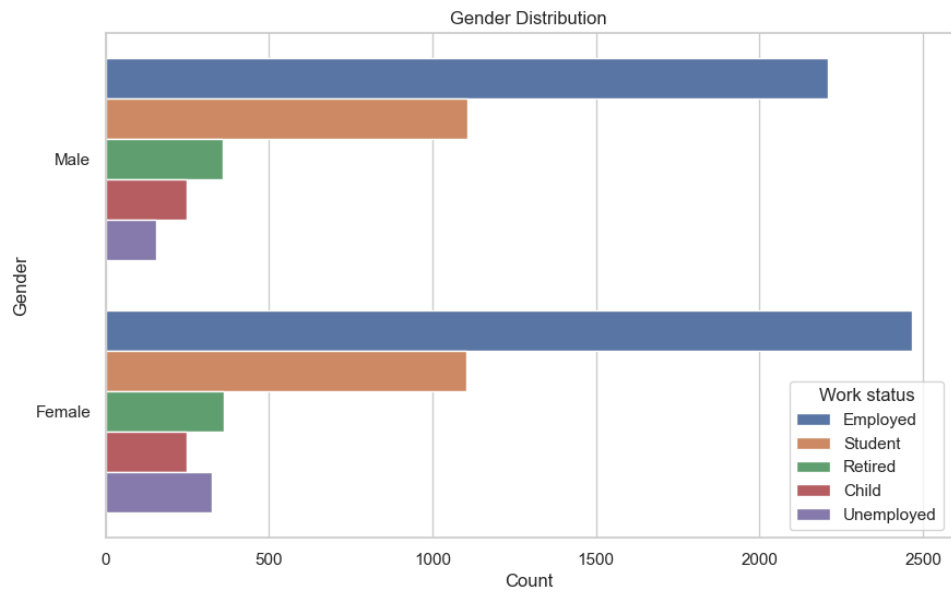
	House Number	Age	Birth_Year	Occupancy	Annual Salary
count	8577.000000	8577.000000	8577.000000	2991.0	5399.000000
mean	37.082896	35.593098	1845.406902	2.916416	31995.943167
std	44.790872	21.376178	21.376178	1.587904	7660.018877
min	1.000000	0.000000	1774.000000	1.0	18148.000000
25%	9.000000	18.000000	1831.000000	2.0	27576.000000
50%	21.000000	35.000000	1846.000000	3.0	35955.110000
75%	42.000000	50.000000	1863.000000	4.0	35955.110000
max	220.000000	107.000000	1881.000000	17.0	50682.880000

Summary statistics shows a population that judged average in income with her mean salary of £31,995.95 falling few thousands below the national average salary of £34,983 at the end of 2023 (www.ons.gov.uk, 2024).

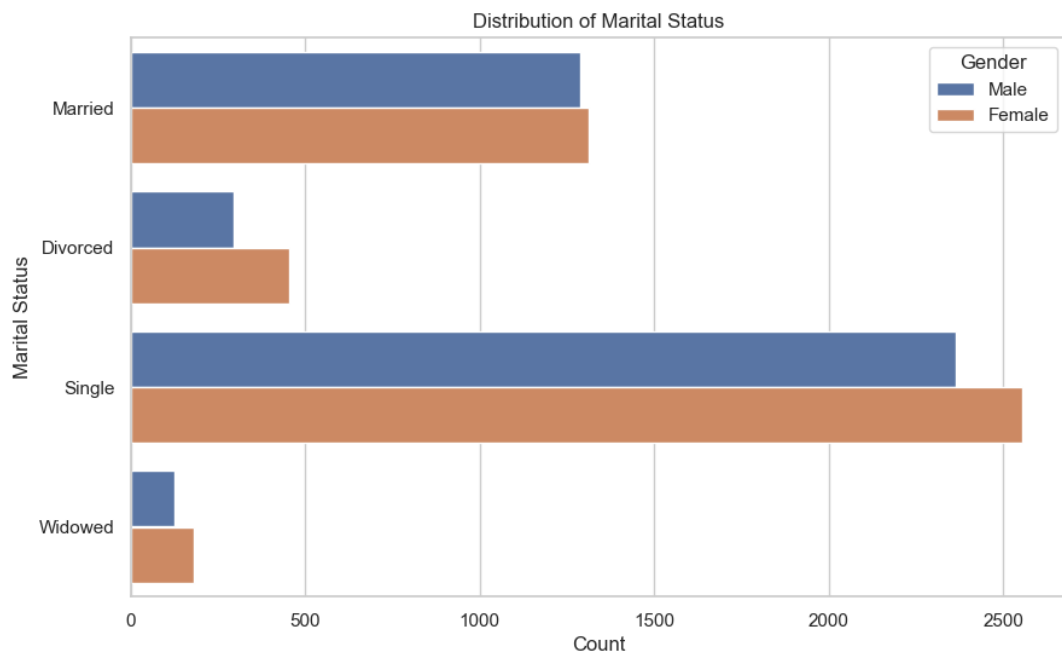


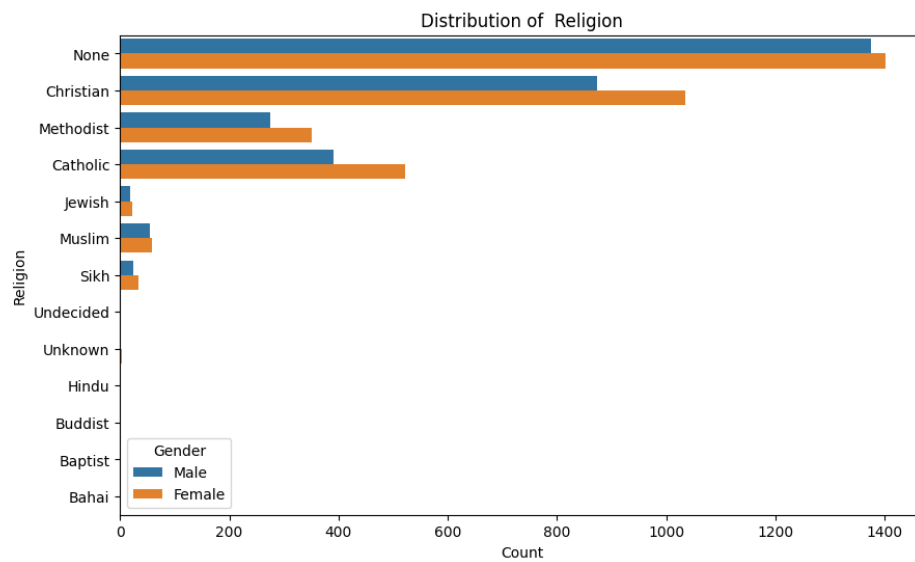
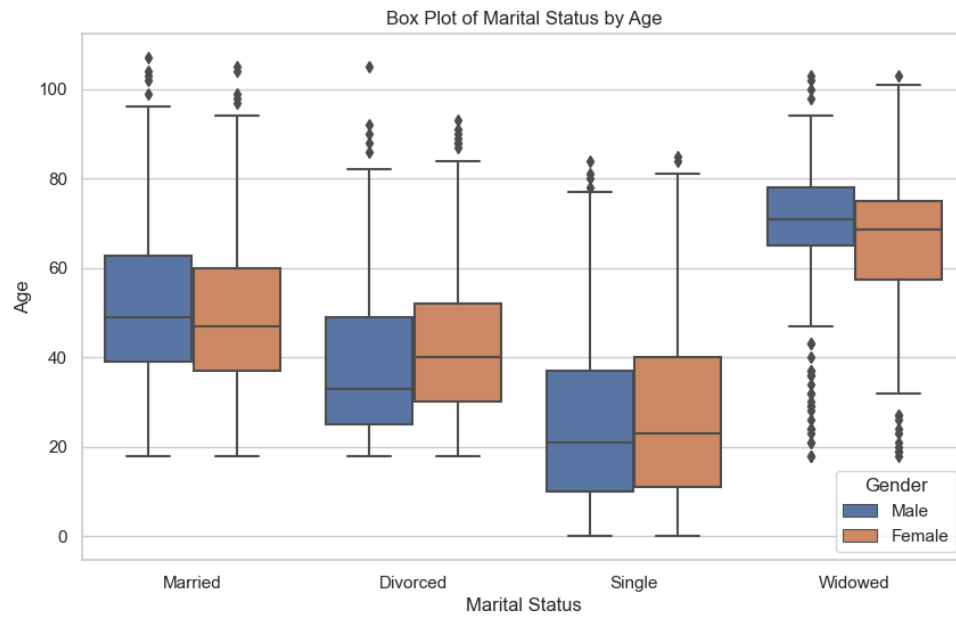
- From the age pyramid, the dominant age in this town is between 30 and 49 followed by 10 to 29 and then 0 to 9. People aged 60 and above are lesser. It shows that it has more middle-aged people who are in the labour force.
- A total of 4674 (54.94%) inhabitants of this town are employed, and the population of the female population is seen topping both the employed and unemployed groups.





From the gender distribution, it is evident that there are more females than males in the population.





Analysis

Population Growth

Growth rate = birthrate / 1000. Growth rate is less than 1%. Birth rate of the last ten years is 12.93 and was calculated by dividing the annual birth by 1000. Estimated death rate for the last ten years of 12.912% was calculated by subtracting growth rate from birth rate.

Occupancy

```
df['Occupancy'].value_counts()
```

```
2      783
3      665
1      575
4      512
5      348
6       58
7       21
10       9
11       8
12       5
8         5
9         1
17        1
Name: Occupancy, dtype: int64
```

The town has a mean occupancy of 3. Only 50 houses have occupancy above 6. The town is not growing and does not have need for high density buildings no need for high density building.

Religion

The number of populations with None as religion does not show enough interest in the population. No one group of those practicing religion is enough to demand urgent attention.

Commuters

2145 earn below the national average salary. There is also have 2210 students in the population. These represents 50.78% of the population and will constitute the bulk of commuters.

Recommendation

1. This town needs a train station to serve needs of the commuting population is not less than half of the total population. More inhabitants who own car will also enjoy the use of the train stations as some do not like to drive always.
2. The growth rate of the town, which has been going the downward trend in the last ten years suggests that the Council should urgently focus on attracting people to the town and offer child support services to encourage childbearing. The council should come up with child support programmes like in Sweden, Canada.

3. While train station contributes to serving commuters, it will go a long way in increasing living conditions of the town to attract new inhabitants and also make the students population like to remain in the town.
4. In summary, a child support program to encourage childbearing for single or married will help the growth rate. Train station contributes to amenities in the town that will attract new dwellers.
5. With more population, the Council will have more funds through tax to run its programme. Although the age range above 70 is 7.45%, this increase in the next 10 year and will be challenging from next 20 years and younger population are expected to fill the gap of working to support the age range of 30 – 49 when they grow old.

References

Castel, F., (2018) Drivers of migration: why do people move? *Journal of Travel Medicine*, Volume 25, Issue 1, , <https://doi.org/10.1093/jtm/tay040>

Vij, A., Ardeshiri, A., Li, T., Beer, A. and Crommelin, L. (2022) Understanding what attracts new residents to smaller cities, AHURI Final Report No. 375, Australian Housing and Urban Research Institute Limited, Melbourne, <https://www.ahuri.edu.au/research/final-reports/375>, doi: 10.18408/ahuri3226201.

Local Government Association Local Government(2013) [councillors-guide-procure-4b2.pdf \(local.gov.uk\)](#)

[7 Project Selection Methods That Will Help You Succeed \(emeritus.org\)](#)

Schwartz, G., Fouad, F., Hansen, T. (2020) Well Spent Genevieve Verdier
Accessed <https://doi.org/10.5089/9781513511818.071>

[World-Bank-Infrastructure-Prioritization-Framework-2016.pdf \(ukniat.ng\)](#)