Census Project Report

This report reviews and analyses the mock census data of a moderately sized town to make decisions on what to do with unoccupied plot of land and possibly support further investment. In order to ensure accurate statistics of the population for better planning, policies implementation and fund allocation where required the mock census data was cleaned to correct data errors and missing records accordingly.

Subsequent sections of the report will show the major analytical approaches implemented to support recommendations provided. This includes, an overview of the town's population demographics, followed by detailed analysis of the town's predicted population growth, employment trends, occupancy rates and commuters.

Data Cleaning

It was done to correct the data errors, solve inconsistencies, anomalies, incomplete and missing records; a detailed statement of all cleaning carried out can be seen in the corresponding Jupyter Notebook.

The first column in the raw dataset which is 'Unnamed:0' was dropped since it prepends no significant input to the project in view.

Blank data were imputed by inferring information from an individual's household, occupation, marital status, age distribution and common denominator (mostly in the cases of missing age, marital status, occupation). Records that could not be located with referential information were converted to 'None' or 'Unknown' (especially in the cases of surname, relationship to head of house, infirmity and religion).

Sith in Religion column is not a religious group. It is an ancient order of Force-wielders that has evolved over thousand years according to Darth Sidious (2012 p.5). Sith was changed and classified as 'None' (because it is the most common category under religion). Nope in religion category was also changed to Atheist since she does not believe in God. According to Peter Boghossian of a Manual for Creating Atheists:

'' This confusion is understandable given that the word "theist" is contained in the word "atheist." It is thus natural to assume a type of parallelism between the two words. Many of the faithful imagine that just as a theist firmly believes in God, an a-theist firmly disbelieves in God. This definitional and conceptual confusion needs to be clarified'' (Boghossian, 2013, p.39)

In marital status, people below 18 had NaN as a marital status which was replaced with Minor. An exception has been made for those 16 or above, as it is legal to marry with a parental consent (Marriage Act, 1949: s3). Though a widower aged 18 was observed. He was changed to single marital status.

Religion for minors (under 18) were primarily NaN entries excluding 2548 entries who were above 18 including initial Sith and one missing religion respectively. All the entries bearing NaN were converted to None along with the missing religion. The implication of this is that it will be difficult to analyse

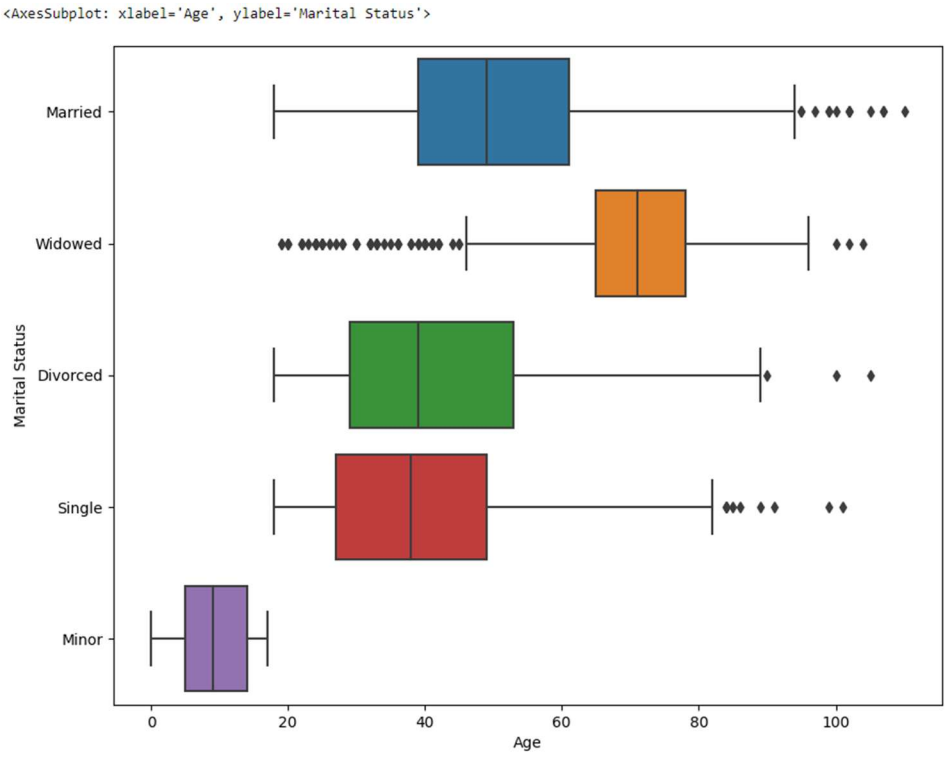
religion transference between the parents and children. Withstanding that majority of the children come from Christian homes further scaling could be done to predict their respective religious groups in near future putting some external factors into consideration.

Individuals with abbreviated letters and case sensitive words were classified and imputed under their respective categories (in the case of gender and marital status). Also, persons with lettered ages were converted to figures in their required datatype- integer.

Table below shows the marital status grouped by age. We could see some outliers across that dataset especially cases of divorcee at age 105, single at 101. Though widow at 104 is quite common but there could be some rare cases of divorce and cohabitation in the aging population. This was indicated in the relationship to head of house where some single aging male partners above 65 years cohabit.

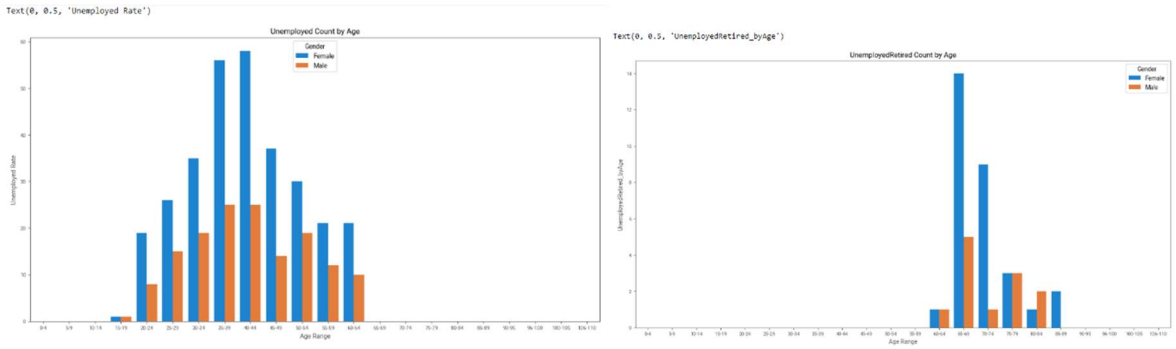
Marital Status	count	mean	std	min	25%	50%	75%	max
Divorced	676.0	42.498521	17.283132	18.0	29.0	39.0	53.0	105.0
Married	2142.0	50.923436	16.720236	18.0	39.0	49.0	61.0	110.0
Minor	1633.0	9.041029	5.111037	0.0	5.0	9.0	14.0	17.0
Single	2440.0	38.584016	13.939081	18.0	27.0	38.0	49.0	101.0
Widowed	369.0	66.943089	17.479787	19.0	65.0	71.0	78.0	104.0

Likelihood fo higher STD for the divorced. Women also report the highest rate of sexual activity and vaginal intercourse at ages 25-29. (Research Gate, 2010).



The above boxplot shows numerous outliers, it is most likely for aged people to become widow or single at old age. Except the student who was already a window at 18 which was imputed as single given justification that he is a student living under his parents without job.

Evident information which outlined Occupation as 'Unemployed' for individuals below retirement age were captured with 'unemployed_population10' whereas those who are retired and unemployed from 65 years above were cascaded to 'unemployedRetired_population10'. Despite the fact that one can be unemployed before retirement age, but it should not be the main focus now as those beyond the retirement range would not usually be eligible for work (Gov.uk).

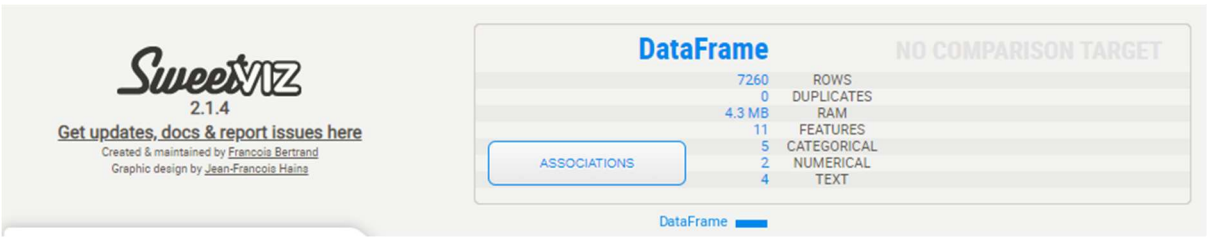


Population Demographics

Having cleaned the given data, the prepared sample of data will have the following information shown below.

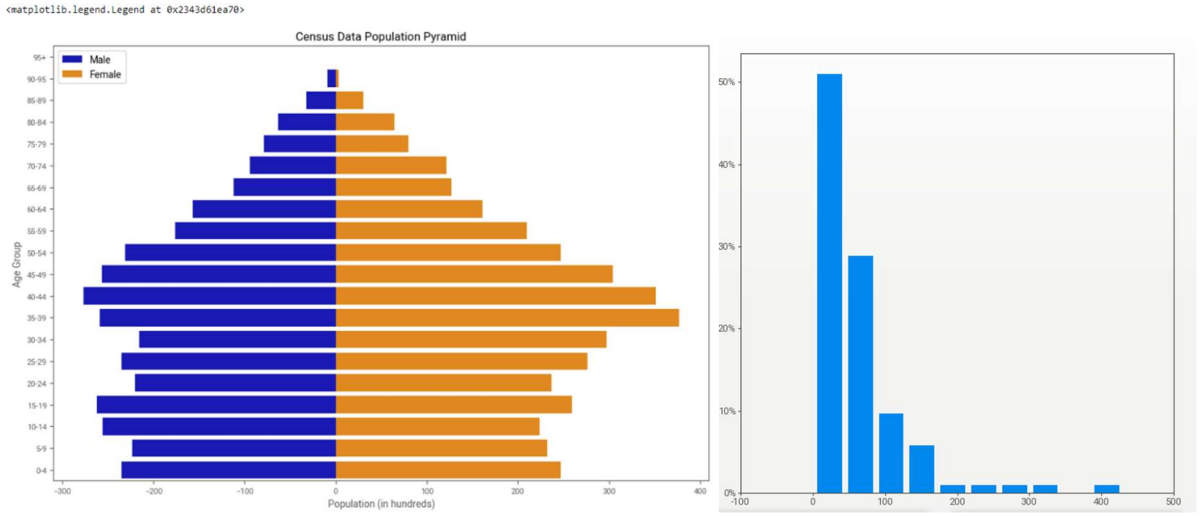
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7260 entries, 0 to 7259
Data columns (total 12 columns):
 #   Column                        Non-Null Count  Dtype
---  ------                        --------------  -----
 0   House Number                  7260 non-null   int64
 1   Street                        7260 non-null   object
 2   First Name                    7260 non-null   object
 3   Surname                       7260 non-null   object
 4   Age                           7260 non-null   int32
 5   Relationship to Head of House 7260 non-null   object
 6   Marital Status                7260 non-null   object
 7   Gender                        7260 non-null   object
 8   Occupation                    7260 non-null   object
 9   Infirmity                     7260 non-null   object
 10  Religion                      7260 non-null   object
 11  Age Range                     7260 non-null   category
dtypes: category(1), int32(1), int64(1), object(9)
memory usage: 603.5+ KB
```

For detailed analysis, Age Range was added to the dataset: This is done by classifying the ages in differences of 5 across the town.



Demographics

In the population pyramid, the structure of the population shows a slightly lower number of young people compared to middle-aged (5-9), indicating a decrease in the number of births within the last decade. They also tend to live well into old age, especially the men. Statistics have it that the first 60 house numbers across the respective Streets have more occupation counts compared to the rest of the streets except for retired unemployed persons.

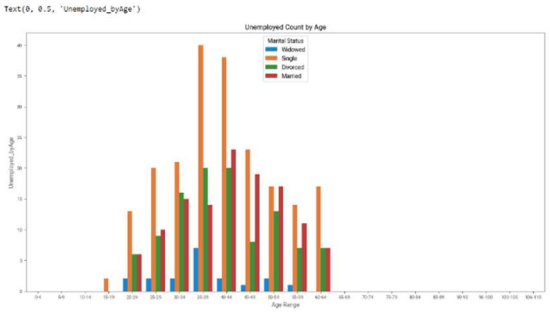


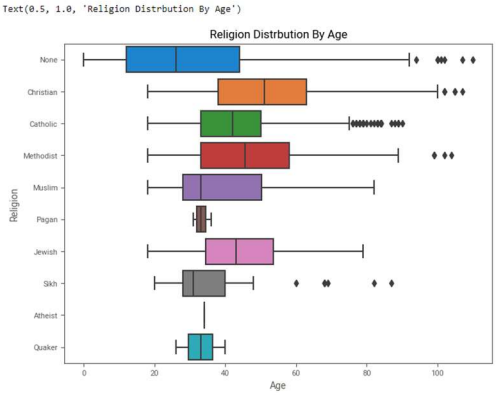
Occupancy shows that there are many single occupants across the Houses especially House numbers 40-44 with a total of 846 counts across the dataset. Relatively less than 1% across the population has infirmity with mental disability more in women (widow and divorced).

Number of Occupants	count	mean	std	min	25%	50%	75%	max
1	846.0	43.955083	49.405870	1.0	10.00	24.0	54.00	220.0
2	827.0	38.307134	44.928892	1.0	9.00	22.0	44.00	221.0
3	446.0	29.831839	38.560980	1.0	8.00	17.0	31.75	203.0
4	347.0	27.881844	35.722200	1.0	7.00	16.0	33.50	191.0
5	267.0	25.411985	32.624008	1.0	8.00	16.0	28.00	188.0
6	47.0	20.787234	24.668177	2.0	6.00	17.0	26.50	163.0
7	8.0	13.750000	13.231456	3.0	3.75	11.5	16.00	43.0
8	7.0	10.000000	10.000000	1.0	3.00	7.0	13.50	29.0
9	4.0	17.000000	10.832051	9.0	12.00	13.0	18.00	33.0
10	9.0	11.333333	10.074721	1.0	4.00	9.0	12.00	31.0
11	3.0	1.333333	0.577350	1.0	1.00	1.0	1.50	2.0
12	6.0	8.666667	12.176480	1.0	1.75	5.0	6.75	33.0
15	1.0	7.000000	NaN	7.0	7.00	7.0	7.00	7.0
18	1.0	12.000000	NaN	12.0	12.00	12.0	12.00	12.0
20	1.0	1.000000	NaN	1.0	1.00	1.0	1.00	1.0
21	1.0	1.000000	NaN	1.0	1.00	1.0	1.00	1.0

TOP CATEGORIES		
None	7,220	>99%
Physical Disability	12	<1%
Deaf	8	<1%
Blind	7	<1%
Disabled	5	<1%
Unknown Infection	5	<1%
Mental Disability	3	<1%
ALL	7,260	100%

Further detailed analysis shows 87.8% of the town under the age of 65 years are employed excluding children, students and university students while 6% are unemployed under the age of 65 years. Most of the unemployed population are single.





Common Values

Value	Count	Frequency (%)
None	4181	57.6%
Christian	1691	23.3%
Catholic	752	10.4%
Methodist	486	6.7%
Muslim	84	1.2%
Sikh	41	0.6%
Jewish	19	0.3%
Pagan	3	< 0.1%
Quaker	2	< 0.1%
Atheist	1	< 0.1%

It is observed that across the population only one person does not believe in God.

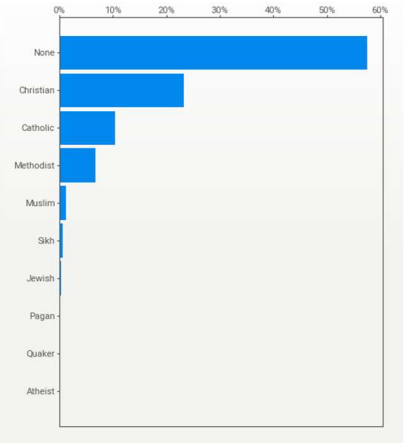
Religion	count	mean	std	min	25%	50%	75%	max
Atheist	1.0	34.000000	NaN	34.0	34.0	34.0	34.00	34.0
Catholic	59.0	42.372881	14.292388	20.0	32.0	41.0	49.00	84.0
Christian	128.0	49.296875	18.811271	18.0	35.0	49.5	63.00	100.0
Jewish	3.0	50.666667	30.730007	18.0	36.5	55.0	67.00	79.0
Methodist	45.0	40.777778	14.355691	19.0	32.0	39.0	45.00	88.0
Muslim	5.0	37.200000	25.508822	19.0	24.0	30.0	31.00	82.0
None	177.0	38.649718	13.999651	18.0	28.0	37.0	46.00	89.0
Sikh	4.0	29.750000	3.593976	25.0	28.0	30.5	32.25	33.0

Marital Status

Common Values

Value	Count	Frequency (%)
Single	2440	33.6%
Married	2142	29.5%
Minor	1633	22.5%
Divorced	676	9.3%
Widowed	369	5.1%

Religion



Infirmity

Common Values

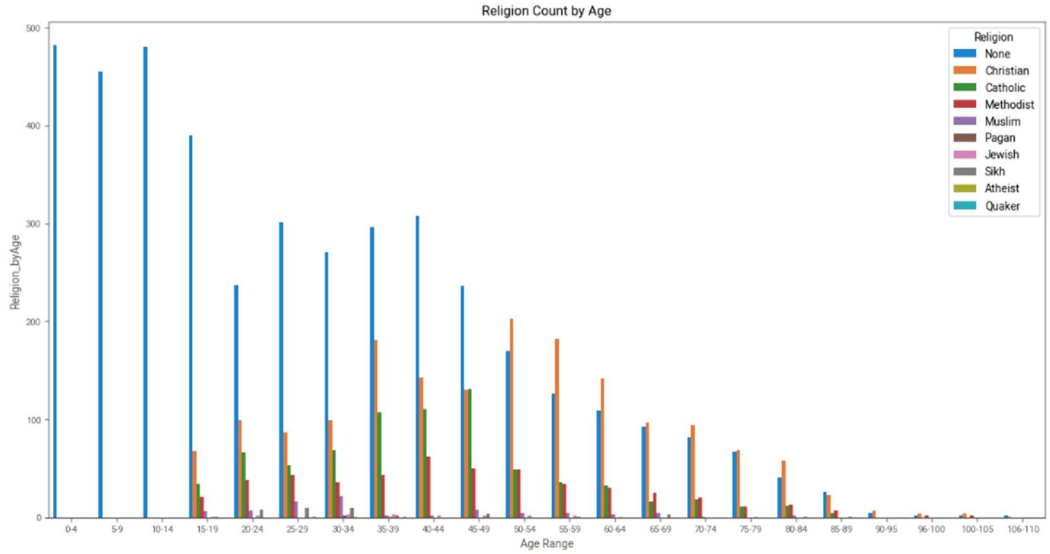
Value	Count	Frequency (%)
None	7220	99.4%
Physical Disability	12	0.2%
Deaf	8	0.1%
Blind	7	0.1%
Disabled	5	0.1%
Unknown Infection	5	0.1%
Mental Disability	3	< 0.1%

Detailed Analysis

Religion and Infirmity

As mentioned earlier, Infirmity comprised of less than 1% of the population. It is not necessary to consider it extensively in the recommendations. There are some observable interests in the religion especially None which are mainly under 18. It could be inferred that most of them come from a Christian or catholic home especially the cases of Lucas and Lewis families respectively. There is a high tendency for religious transference with time.

```
Text(0, 0.5, 'Religion_byAge')
```

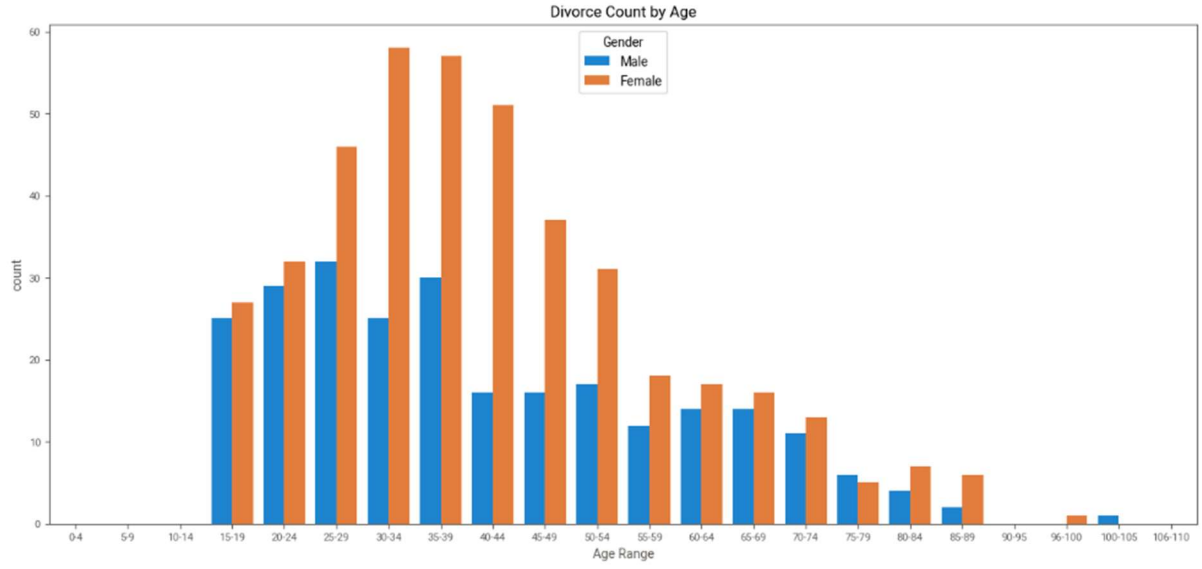


Divorce and Marriage

Results of analysis show that divorce occurs from young age of 15-19 up to old age of 106-110. Divorce count split by gender demonstrates that there are more female divorcees than the male counterparts. It could be deduced that there is a potential male divorcees drift from the town.

Calculating the crude divorce rates, marriage rates and divorce rates; the statistics of divorced count must be done with reference to divorced women. Marriage is calculated by counting the number of married individuals and dividing the sum by 2. In population of 7260 there are a total count of 422 female divorcees and 1071 active couple indicating that the divorce rate is high. In any given sample 1000 population within the town there are 58 and 148 female divorcees and couples respectively.

```
Text(0.5, 1.0, 'Divorce Count by Age')
```



Birth and Death Rate

Mathematical calculation of birth rates shows that the population is not growing though there are likelihood of more ageing population. The crude birth rate and death rate are calculated as:

Crude Birth Rate = (Number of annual births / Number of total population) x 1,000.

The present crude birth rate is approximately 7 births per thousand. It could be assumed that there is high usage of contraceptives within the town.

Crude Death Rate = (Number of annual Deaths / Number of total population) x 1,000.

The death rate is calculated by estimating deaths by difference in age-bands for those above 55. According to the age range pyramid, those who are above 55 are most likely to retire and differences in these age ranges have more death trends. The summation of the differences for those above 55 and division by 5 (based on age range), gives the death rate of 15 per thousand which is double the birth rate.

	Age Range	Total Population	Age Difference
0	106-110	3	-5.0
1	100-105	8	0.0
2	96-100	8	-4.0
3	90-95	12	-50.0
4	85-89	62	-65.0
5	80-84	127	-32.0
6	75-79	159	-57.0
7	70-74	216	-23.0
8	65-69	239	-79.0
9	60-64	318	-68.0
10	55-59	386	-92.0
11	50-54	478	-83.0
12	45-49	561	-67.0
13	40-44	628	-8.0
14	35-39	636	123.0
15	30-34	513	2.0
16	25-29	511	54.0
17	20-24	457	-64.0
18	15-19	521	41.0
19	10-14	480	25.0
20	5-9	455	-27.0
21	0-4	482	NaN

Crude Birth Rate per 1000

The birth rate was calculated by dividing the number of children born annually by the total population and multiplying the result by 1000.

```
age_0 = sum(df_census10['Age'] == 0)

i = 1000/ len(df_census10) #this will be used as the multiplication constant when there is
                #need to express any number per thousand of the population

Crude_birth_rate = age_0 * i
Crude_birth_rate
```

```
6.887052341597796
```

The birth rate is approximately 7 in a thousand population which is quite low

Progressive Birth Rate

This will take into account the number of children delivered annually by the number of population within the age range of 5.

```
age_5 = sum(df_census10['Age'] == 4)
pop_35_39 = 636
progres_birt = age_5 * i
progres_birt
```

```
11.84573002754821
```

Migration

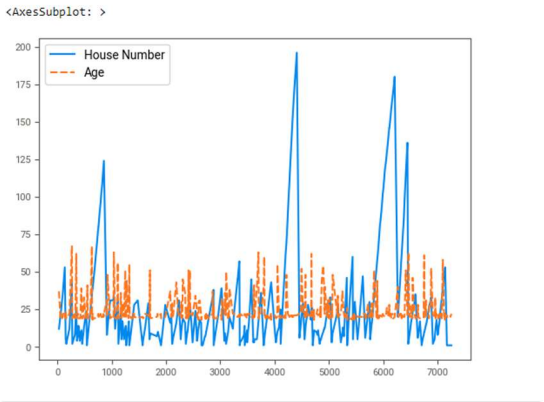
For Immigration **t**he following relationships to head of house were considered: None, Visitor, Lodger, Adopted Daughter & Adopted Son respectively. This is including those who are divorced and single probably have left their spouse from another city. A total count of 79 immigrants per thousand population was observed. However, emigration details are gotten from the difference in male and female divorcees within the population. Implementing the above method emigration from the given dataset is approximately 24 per thousand.

Employment and Commuters

Commuters are picked based on the following assumptions:

- Individual who is identified with a university since there is none within the town (PhD Student, University Student, Professor Emeritus, Associate Professor, Lecturers including further and higher education).
- Oceanographer bearing in mind that there is no water related information about the town.
- Solicitors, Barrister, Barista need to commute to their respective courts within and outside the town.

- Researchers were assumed to be affiliated to a university regardless of their disciplines.



Occupation	count	mean	std	min	25%	50%	75%	max
Associate Professor	2.0	30.000000	8.485281	24.0	27.00	30.0	33.00	36.0
Barista	4.0	31.000000	8.755950	20.0	26.00	32.5	37.50	39.0
Barrister	5.0	46.600000	8.961027	37.0	37.00	51.0	52.00	56.0
Further education lecturer	7.0	45.571429	14.257830	26.0	36.00	45.0	56.50	63.0
Higher education lecturer	7.0	31.142857	3.078342	28.0	28.50	31.0	33.00	36.0
Lecturer, further education	5.0	45.800000	12.194261	36.0	40.00	43.0	43.00	67.0
Oceanographer	8.0	37.875000	10.802612	27.0	29.50	33.0	50.25	51.0
PhD Student	13.0	27.384615	2.631174	23.0	26.00	28.0	29.00	32.0
Professor Emeritus	4.0	46.000000	13.140269	31.0	40.75	45.0	50.25	63.0
Research scientist (maths)	7.0	43.142857	9.026416	29.0	39.00	40.0	50.00	55.0
Research scientist (medical)	10.0	39.200000	12.407883	24.0	26.50	42.0	47.75	58.0
Ship broker	10.0	41.900000	13.486207	26.0	31.75	36.5	53.25	62.0
Solicitor	3.0	44.666667	16.802778	30.0	35.50	41.0	52.00	63.0
Solicitor, Scotland	12.0	45.583333	12.464191	25.0	36.75	46.0	54.25	67.0
University Student	404.0	20.299505	1.218916	18.0	19.00	20.0	21.00	24.0

It is on record that the university student commute more compared to other road users with a minimum of 501 individuals using the roads on daily basis. Bearing in mind that other professionals are likely to use the motorways periodically though this partly will influence our recommendations.

As stated earlier, analysis shows 87.8% of the working-age population (under the age of 65 years) are employed excluding children, students and university students while 6% are unemployed under the age of 65 years.

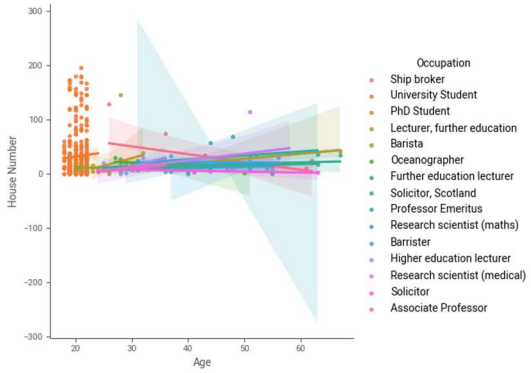
Occupancy Rates

The median occupancy rate of the town is 2 while for the house number is 19. This does not actually mean that every home has a median two bedrooms, but it acts as a factor to reiterate over-occupancy. Using this median, there are 827 homes over-occupied – increasing the median to three shows 446 homes over occupied. Incidentally, there is a tendency for household occupancy to be higher as the median occupancy drops.

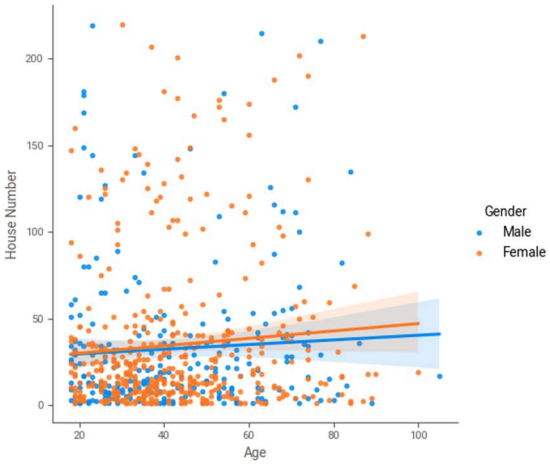
Extremely high occupancy (10+) could be attributed to large families living together and students, though these could also be chunks of apartments.

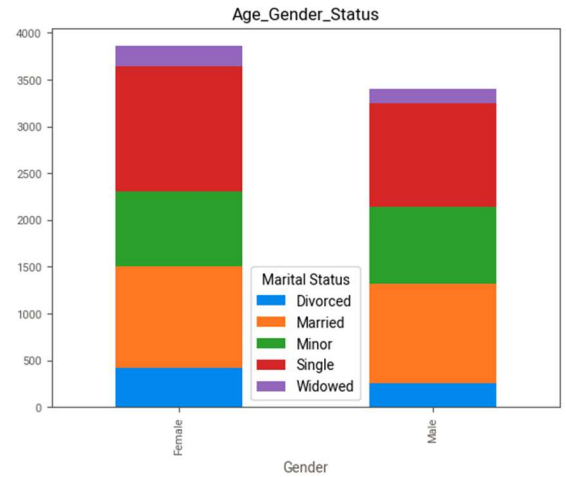
Number of Occupants	count	mean	std	min	25%	50%	75%	max
1	846.0	43.955083	49.405870	1.0	10.00	24.0	54.00	220.0
2	827.0	38.307134	44.928892	1.0	9.00	22.0	44.00	221.0
3	446.0	29.831839	38.560980	1.0	8.00	17.0	31.75	203.0
4	347.0	27.881844	35.722200	1.0	7.00	16.0	33.50	191.0
5	267.0	25.411985	32.624008	1.0	8.00	16.0	28.00	188.0
6	47.0	20.787234	24.668177	2.0	6.00	17.0	26.50	163.0
7	8.0	13.750000	13.231456	3.0	3.75	11.5	16.00	43.0
8	7.0	10.000000	10.000000	1.0	3.00	7.0	13.50	29.0
9	4.0	17.000000	10.832051	9.0	12.00	13.0	18.00	33.0
10	9.0	11.333333	10.074721	1.0	4.00	9.0	12.00	31.0
11	3.0	1.333333	0.577350	1.0	1.00	1.0	1.50	2.0
12	6.0	8.666667	12.176480	1.0	1.75	5.0	6.75	33.0
15	1.0	7.000000	NaN	7.0	7.00	7.0	7.00	7.0
18	1.0	12.000000	NaN	12.0	12.00	12.0	12.00	12.0
20	1.0	1.000000	NaN	1.0	1.00	1.0	1.00	1.0
21	1.0	1.000000	NaN	1.0	1.00	1.0	1.00	1.0

<seaborn.axisgrid.FacetGrid at 0x2373a152080>

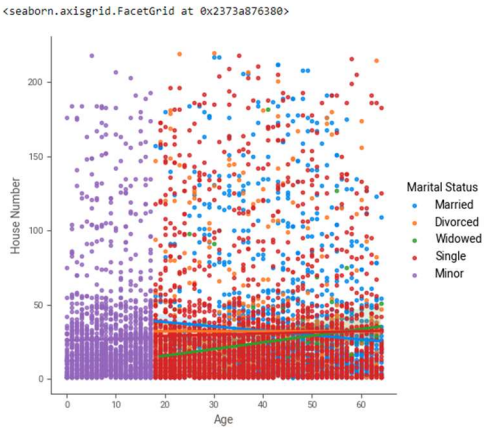


<seaborn.axisgrid.FacetGrid at 0x2373a465150>





<seaborn.axisgrid.FacetGrid at 0x2373a876380>



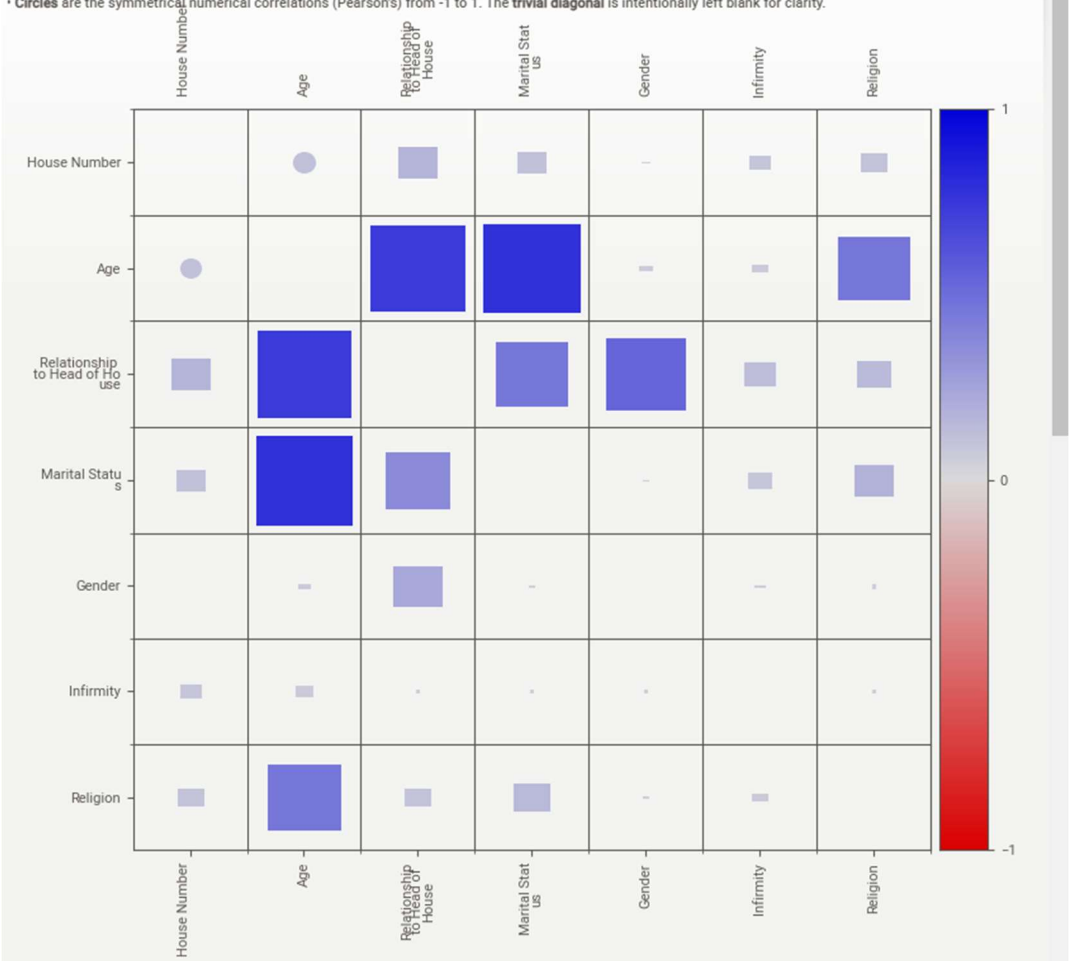
This shows that there are a lot of under 18 and single active working-age population within the town.

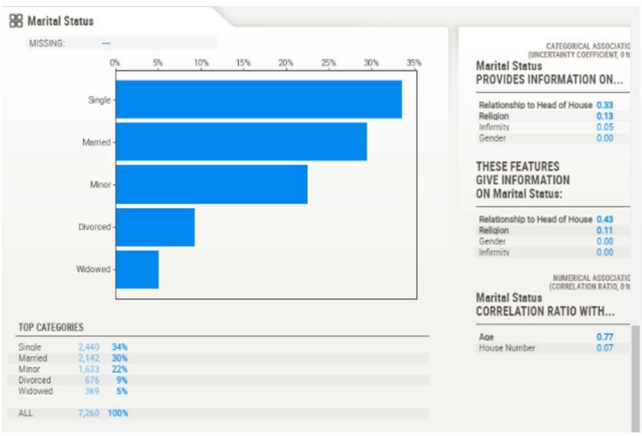
Associations

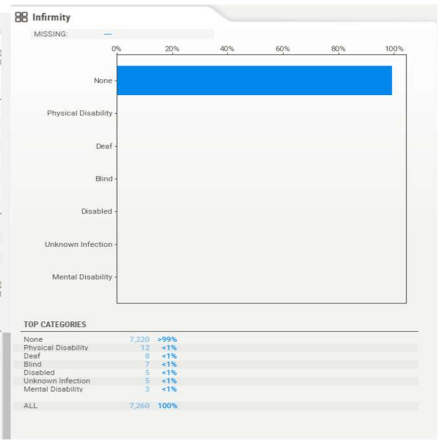
[Only including dataset "DataFrame"]

▪ **Squares** are categorical associations (uncertainty coefficient & correlation ratio) from 0 to 1. The uncertainty coefficient is **assymmetrical**, (i.e. ROW LABEL values indicate how much they PROVIDE INFORMATION to each LABEL at the TOP).

• **Circles** are the symmetrical numerical correlations (Pearson's) from -1 to 1. The **trivial diagonal** is intentionally left blank for clarity.







Recommendations

Statistically, taking into consideration our birth rates and death rates as demonstrated above shows that the town population is not growing thus there is no feasible significant expansion in population. As a result of this declining birth rates, High-density housing should not be encouraged.
Since there is a tendency for household occupancy to be higher as the median occupancy drops. This suggests that there is relative demand for Low density housing where there is need for larger family housing. This would benefit more lodgers, visitors, adopted children and unidentified individuals who are neither related nor unrelated to the head of house('None') moving to the town. Also, for divorcees who are living with their children.

There is a need to construct a centralized railway station (within the town). This will decongest the motorways especially for the students and other commuters who cannot afford to spend money for motor transportation but prefer to personal bikes to commute. This will offer comfort, fast long distant services to the diverse professionals residing within the town who might need it periodically.

Since there is high tendency of religious transference between parents and children as discovered in some Christian households. Though this is not quite pronounced across the dataset, but it would be necessary to put up a second church preferably catholic church because of its trajectory across the age range.

Siting an Emergency medical building for minor injuries is advisable taking note of the dense adolescent populace within the age range 10 -15 who are more prone to unintentional injuries (Grossman 2000 p.23). This should be considered a priority because its first line of call is to save lives.
It is evident that there are many unemployed females within the 6% unemployed population in their active working-age distribution. There is need for training and skill acquisition.
Though there are declining birth rates, but it could be figured out that the aging population reasonably stay long. Also, from analyses a lot of unemployed retired individuals were observed reasonably stable with little or no infirmity. There is exigency to allocate money for end-of-life care and post retirement jobs.
In school-aged children education, it is not needful to increase spending for schooling since there is no evidence of growing population within their age range as shown on population pyramid.
In conclusion, bearing in mind that the aging population will only expand and will require more care and post retirement plans. It is very important to invest in old age care which will proffer more solution to the unemployed population.

References

Boghossian Peter (2013) A Manual for Creating Atheists. Pitchstone Publishing Durham, NC27705.

Darth Sidious, Book of Sith_ Secrets from the Dark Side [Vault Edition] Copyright © 2012 by Lucasfilm Ltd.

Gov.uk (No Date) Universal Credit Eligibility.

Available Online: https://www.gov.uk/universal-credit/eligibility [Accessed 05/12/2020].

Grossman DC. (2000) The history of injury control and the epidemiology of child and adolescent injuries. Future Child Spring-Summer;10(1):23-52. PMID: 10911687.

Marriage Act (1949) Section 3.

Available online: https://www.legislation.gov.uk/ukpga/Geo6/12-13-14/76/section/3 [Accessed 05/12/2020].