Census Project Report

This report focuses on the census of a moderately sized imaginary town for which recommendations are to be made for investments in future services and find the appropriate use of an unused plot of land for the people of the town. To achieve this objective first the data provided is thoroughly cleaned and corrected data is replaced with the data errors within the records, which will be explained in detail in the first section of this report, also some data categories types have been modified into integers to support better analysis.

Further sections of this report will feature the analysis undertaken specifically to answer the above mentioned questions of usage of plot of land and the recommendations for investment using the town's population demographics , detail analysis of town's predicted population growth including birth rates and immigrations , employment trends and classifications , commuters and occupancy rates .

Data Cleaning

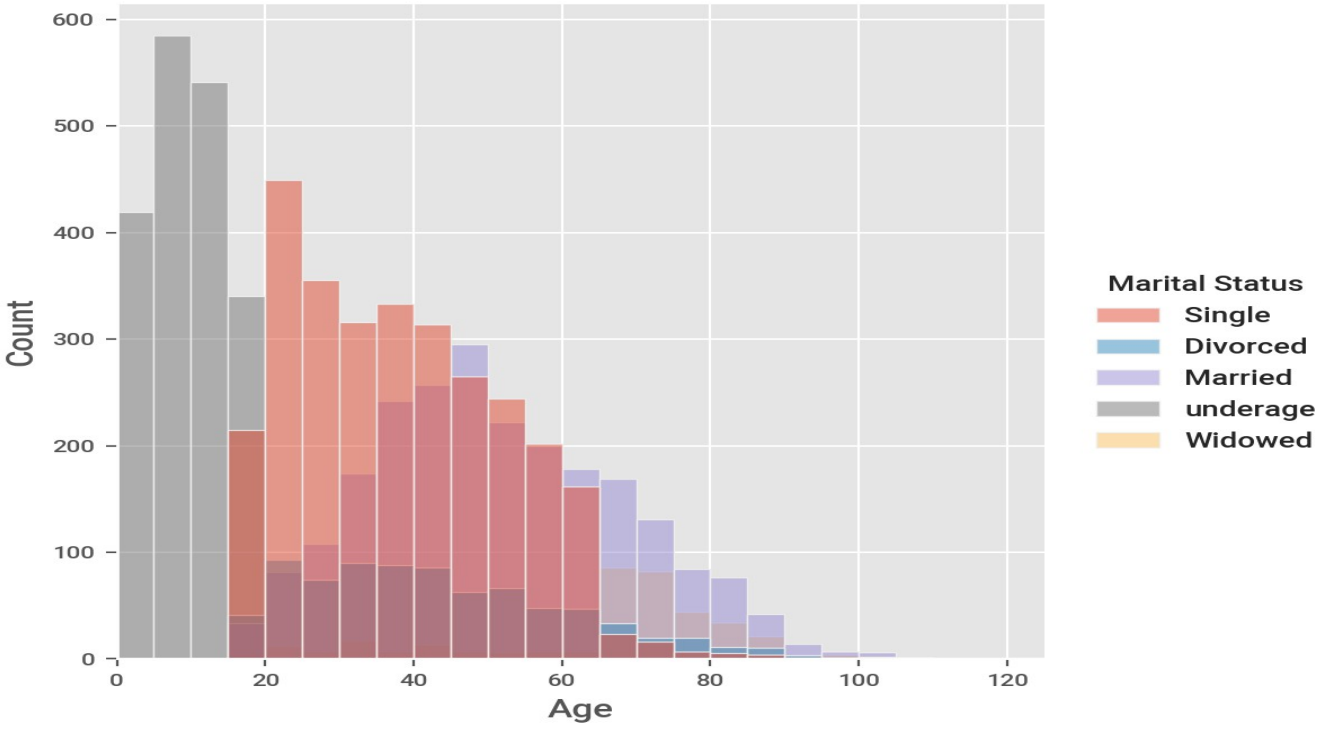
Census data was cleaned to correct the data errors, a full logbook of the data error cleaning can be found in the referenced Jupyter Notebook.

House numbers and Age types were converted from string into integers and in places where alphabetical representation was used were replaced with its integer representations.

Blank data found within the census data were replaced by the person's records using data from their other household members (for surnames, House numbers, Marital Status and Religion in some cases

Religion status for people whom are minors below age 18 were classified in data as NaN entries except a few that were termed as "Married" but were later changed into "Undeclared", there was found to be nearly 25% Nan values for minors in the census data based on that finding the impact of religious transmission from parents to their kids was not possible unless more data was provided. The remaining three religions were converted to "None" – "Sith","Nope","Housekeeper". These entries were deliberately misleading inputs, however Housekeeper can be deemed as individual errors.

All the outliers in the ages column that were exceeding 122 years were capped to 122 , as the longest recorded person to ever lived is 122 years and 164 days old (Jeanne Calment) . The plot below shows an example of the final output of Marital Status and Age distribution with a cap age of 122 and corrected Marital Status types.



As you can see in the graph above the ages have been capped and relevant changes have been made to the Marital Status for underage entries in the census data.

There are Nan values in columns “House Number”, “First Name”, “Marital Status” and “Religion” with total count shown below. The above discussed scheme was used to replace Nan values with appropriate data entries.

```
House Number                      True
Street                           False
First Name                        True
Surname                          False
Age                              False
Relationship to Head of House    False
Marital Status                    True
Gender                           False
Occupation                       False
Infirmity                        False
Religion                          True
dtype: bool
```

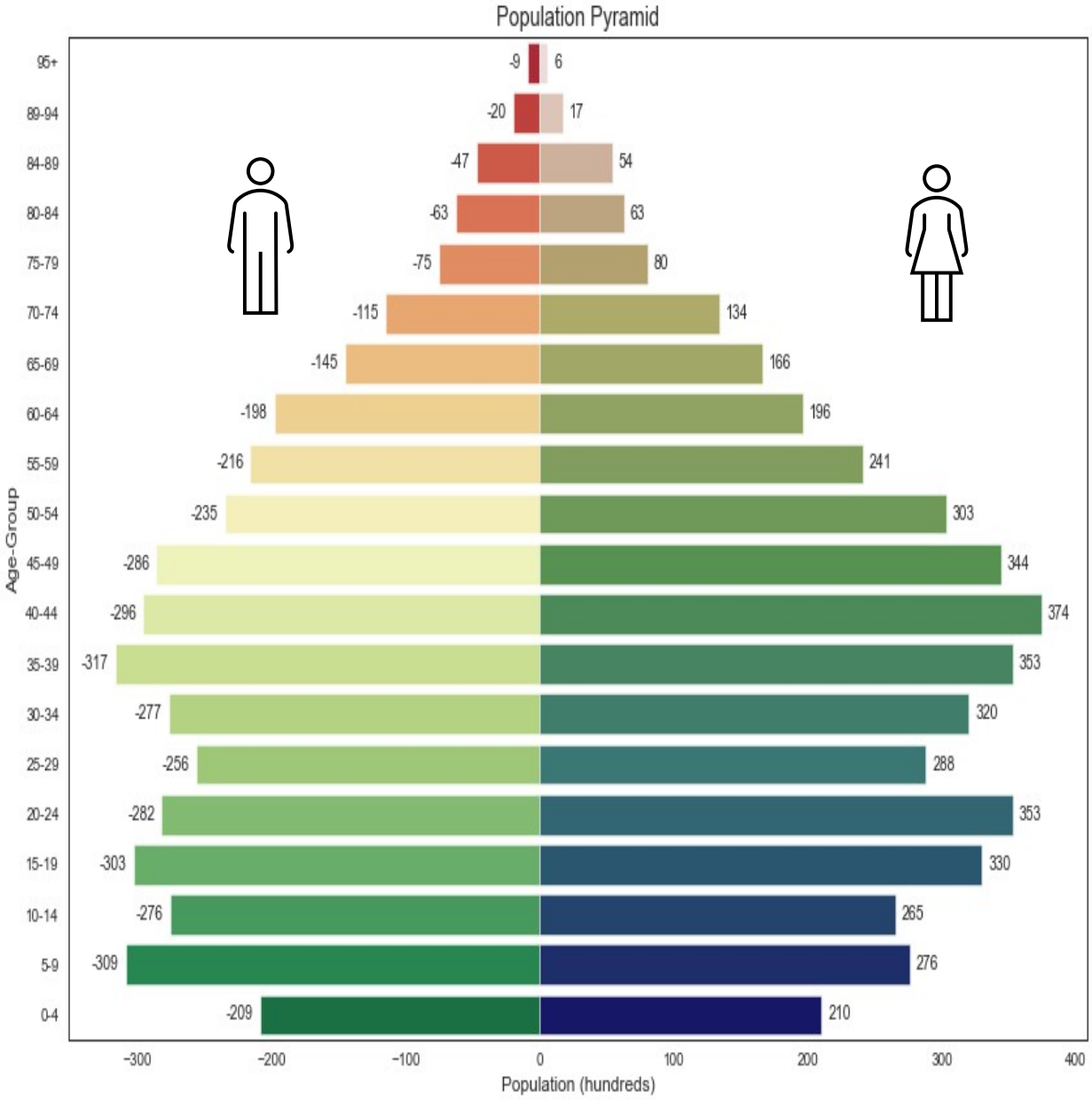
```
House Number                      3
Street                            0
First Name                        2
Surname                           0
Age                               0
Relationship to Head of House     0
Marital Status                 1881
Gender                            0
Occupation                        0
Infirmity                         0
Religion                       1926
dtype: int64
```

Population Demographics

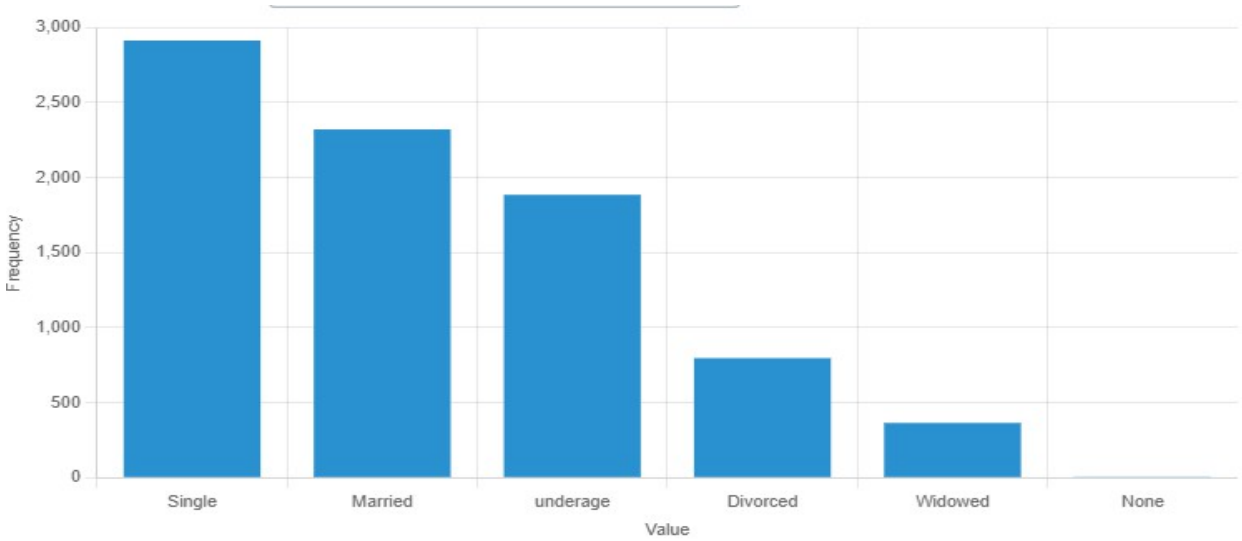
Once data cleaning has been completed the final census data will have the following data types:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8282 entries, 0 to 8281
Data columns (total 11 columns):
 #   Column                        Non-Null Count  Dtype
---  ------                        --------------  -----
 0   House Number                  8282 non-null   int32
 1   Street                        8282 non-null   object
 2   First Name                    8282 non-null   object
 3   Surname                       8282 non-null   object
 4   Age                           8282 non-null   int32
 5   Relationship to Head of House 8282 non-null   object
 6   Marital Status                8282 non-null   object
 7   Gender                        8282 non-null   object
 8   Occupation                    8282 non-null   object
 9   Infirmity                     8282 non-null   object
 10  Religion                      8282 non-null   object
dtypes: int32(2), object(9)
memory usage: 647.2+ KB
```

Examining the population pyramid, the structure of the population shows a lower number of young people compared to middle-aged , especially those aged 0-4 , suggesting a low birth rateThe population also tends to live well into old age, for both male and female. However the low birth rate can be due to the population having more old people than young ones hence explaining the low birth rate .
The population pyramid explains the age group distribution of the population and the number of different groups existing in the town.



Further analysis of the census data shows the the marital status of the majority of the population and that also supports the previous argument of the existence of low birth rate.

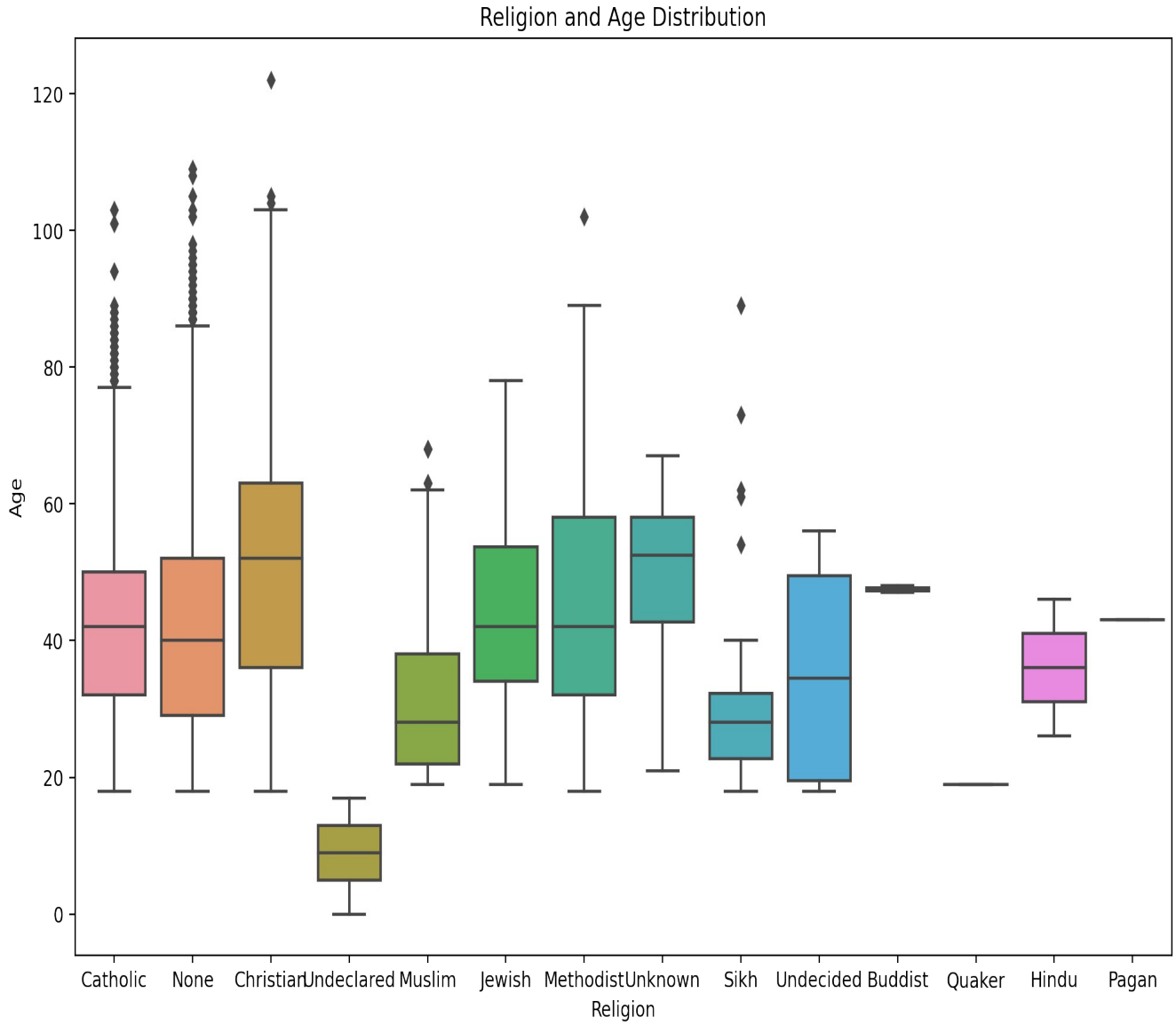


The descriptive analysis of the town population shows that most of the people are Single and then the biggest two marital status that occurs are Married and Underage with a significant small portion Divorced , Widowed and None.

Detailed Analysis

Religion and Age:

The analysis conducted on Religion with the filter of age over the entire population deemed it necessary to highlight that significant religion is Christianity but even bigger portion of population has "None" or hasn't classified their religion including minors and few adults .The plot shown below shows the distribution based on Religion and Age and supports the above mentioned claims. However since christianity is still the dominant religion adding another church based on the growth of the population wont add any more significance , also it is necessary to highlight that the minor religions did grow significantly but not as much needed to make an impact on the final decision making.



Divorce and Marriage

After cleaning the data for census , it was found divorces occured from a young age , starting from 18 years there are 315 male divorcees and 482 female divorces in the census data , femaledivorcees are slightly higher from the male divorcees indicating that male divorcees potentially leave the town more than the females hence defining the difference in the numbers.

Marriage is calculated by counting the number of 'married' individuals and dividing by two. Thus, the divorce to marriage ration is 800 divorces per 1160 marriages which shows increase in crude divorce compared rate (Eurostat,2020) compared to earlier statistic (Eurostat, 2017) .

Birth and Death Rate

Analyzing the birth rate and death rate of the town gives the end result that the town of our census data is experiencing shrinkage or in other words the number of deaths exceeds the number of births in the town. The date rate is calculated by estimating the deaths by difference in age bands 56-60 , 61-65 and 66-70 , 71-75 respectively , by using these two age bands we can find the close to potential death rate of the town . The same method is applied for finding the births in the Town by using the age bands of females from 25-29 , 30-34 respectively . A magnification factor was used (mag) to potray the Birth and Death rates per 100,000 people population , as shown in the code below.

For Deaths rate:

```
people5660=df[(age>=56) & (age<=60)]
people6165=df[(age>=61)&(age<=65)]
deaths=len(people5660)-len(people6165)
tot=deaths*mag
print("Age group 56-60 : %f and Age group 61-65 : %f "% (len(people5660),(len(people6165))))
print("Deaths are in 5 years: %d"%(deaths))
print("death rate of 100K is of 5 years : %f "%(tot))
print("Death Rate per annum per 100,000 people is : %f"%((deaths*mag)/5))
```

```
Age group 56-60 : 443.000000 and Age group 61-65 : 369.000000
Deaths are in 5 years: 74
death rate of 100K is of 5 years : 893.503985
Death Rate per annum per 100,000 people is : 178.700797
```

```
people6670=df[(age>=66) & (age<=70)]
people7175=df[(age>=71) & (age<=75)]
deaths1=len(people6670)-len(people7175)
tot1=deaths1*mag
print("Age group 66-70 : %f and Age group 71-75 : %f "% (len(people6670),(len(people7175))))
print("Deaths are in 5 years: %d"%(deaths1))
print("death rate of 100K is of 5 years : %f "%(tot1))
print("Death Rate per annum per 100,000 people is : %f"%((deaths1*mag)/5))
```

```
Age group 66-70 : 304.000000 and Age group 71-75 : 231.000000
Deaths are in 5 years: 73
death rate of 100K is of 5 years : 881.429606
Death Rate per annum per 100,000 people is : 176.285921
```

For Birth rate:

```
babies=df[(age==0)]
babies
ladiesfilt1=df[(gender=="Female") & (age >= 25) & (age <= 29)]
ladiesfilt1
mag=100000/8282
mag
babiesper100k2529=(len(babies)/len(ladiesfilt1))*mag
babiesper100k2529
babies4=df[(age == 4)]
babies4
ladiesfilt2=df[(gender=="Female") & (age >= 30) & (age <= 34)]
ladiesfilt2
babiesper100k3034=(len(babies4)/len(ladiesfilt2))*mag
babiesper100k3034
rbabies=(babiesper100k2529/babiesper100k3034)*100
tot_bab=babiesper100k2529+babiesper100k3034
print("Total babies born per annum are : %d " % int(tot_bab))
```

```
Total babies born per annum are : 6
```

Migration

After deep diving into the census data of the population it was noticed that there are lodgers and visitors in the population census data , it could be infered that a certain fraction of there are deciding whether to move to the town more permanentaly , the majority of these people are unemployed and they immigrated to the town for job seeking or similar reasons. However few of them visitors are above the retirement age and a significant fraction of them can move to the town permanentaly for settling.
The second thing to analyse was the town has many divorcees , and certain of their partners moved elsewhere cause their records were missing or they changed their surnames. based on the above discussion there are 1292 people per 100,000 moving in the town .

After calculating the immigration and emigration based on the above criteria (including students and unemployed people) the rate of immigration is higher than the emigration of the town.

This brings our population growth of the town more compared to the shrinkage . Population growth per annum on 100K people frame is 859 as shown below.

```
print("\033[1m\"Final Analysis of Dataframe For Growth/Shrinkage on a 100K p
print("Total Deaths : %d "%tot_deaths)
print("Total Births : %d "%tot_bab)
print("Total Divorced moving out : %d "%(tot_divo*mag))
print("Total Immigrants moving in are : %d " %Tot_imm)
print("Total Emigrants moving out are : %d " %Tot_emi)
print("Population growth : %d "%(Total_in-Total_out))
```

```
"Final Analysis of Dataframe For Growth/Shrinkage on a 100K people frame per an
num"

Total Deaths : 354
Total Births : 6
Total Divorced moving out : 24
Total Immigrants moving in are : 1292
Total Emigrants moving out are : 85
Population growth : 859
```

Employment and Commuters

Commuters are identified based on the following methodology:

- Anyone who identifies as a Student
- Anyone who identifies as a University Student
- Anyone who identifies as a Phd Student

Occupations such as teaching (except Higher Education), baristas, retail workers, community roles or food service industry roles were considered non - commuting occupations . Using the above mentioned criteria it was found roughly more than 25% of the total population were commuting or using its services frequently.

Occupancy Rates

The average occupancy rate of the town is 3 , using this it was found there are 1749 people living more than the street average occupation or in other words 1749 people was living in a saturated housing . Although the current residential housing space available (buffer) is 1238 which can cover a major part of the needs of housing if we assume that residential space is available up for rent , However there is still left 511 people needs of housing space in the current framework of the population of the town. So it is recommended to focus on the housing as 511 people still requires it and since it has been established that the population is growing overall after checking the birth and death rates and immigration and emmigration altogether so this will add further load to the current occupancy space available in future (Please check the Jupter Notebook for housing Analysis and Street and Town statistics).

The possible reasons for these saturated housing can be

1. Families have over occupied and have more house members than the town and street average
2. Families can't downsize and are renting rooms to lodgers
3. Families have frequent visitors that stays for a longer duration

Streets	Average House Occupancy	Streets	Average House Occupancy	Streets	Average House Occupancy	Streets	Average House Occupancy
Leeds Lane	3	Scotland Groves	1	Ali Avenue	2	Brick Hills	6
Robertson Road	6	Watson Avenue	2	Rhubarb Road	2	Johnson Light	3
Stanley Road	2	Newry Walks	7	Roberts Avenue	3	Thames Lane	3
Sunderland Bypass	2	Castor Road	3	Richards Drive	2	Fitzgerald Road	1
Walsh Avenue	2	Clarke Avenue	7	Lunar Terrace	2	Jones Plains	3
Milldraw Drive	2	Thompson Corners	2	Briggs Avenue	1	Nottingham Drive	3
Lee Drive	6	Martin Plain	2	Holmes Center	4	Libra Manor	4
Memorial Spur	3	Open Lane	2	Brown Road	3	Norwich Lane	2
Cancer Radial	2	Pilgrim Avenue	2	Badgerdike Street	2	Dragons Lane	2
Easter Unions	2	Anvil Avenue	2	Christmas Street	3	Thistle Ways	2
Kaur Trail	2	Wilson Inlet	1	Canterbury Avenue	3	September Motorway	2
Reindeer Avenue	2	Hall Avenue	3	Phillips Forge	3	Mace Road	3
Brady Dale	3	Libracrib Lane	1	Bank Lane	2	Foster Road	4
Cooper Port	3	Testament Street	2	Coventry Drive	4	Grace Avenue	10
Anvil Heights	1	Ranger Drive	3	Sharpe Parkways	3	Coronationnod Road	4
William Drive	2	Hazel Road	3	Apricot Lane	3	Holyrood Road	2
Dockers Road	3	Benson Road	3	Lisbon Avenue	2	Cardiff Drive	2
Cook Avenue	2	Parry Drive	2	Nottingham Road	3	Kielder Lane	2
Coconut Drive	3	Adams Vista	2	Coventry Avenue	3	Corporationdike Drive	4
Chapman Viaduct	2	Robinson Way	4	Matthews Avenue	3	Bulldog Avenue	4
Thomas Lane	3	Shaw Shore	3	Hussain Street	1	Baxter Road	3
Hangar Orchard	2	Short Mountains	1	Walker Path	4	Mustard View	2

Average Town Occupancy	3.390476 equivalent to 3

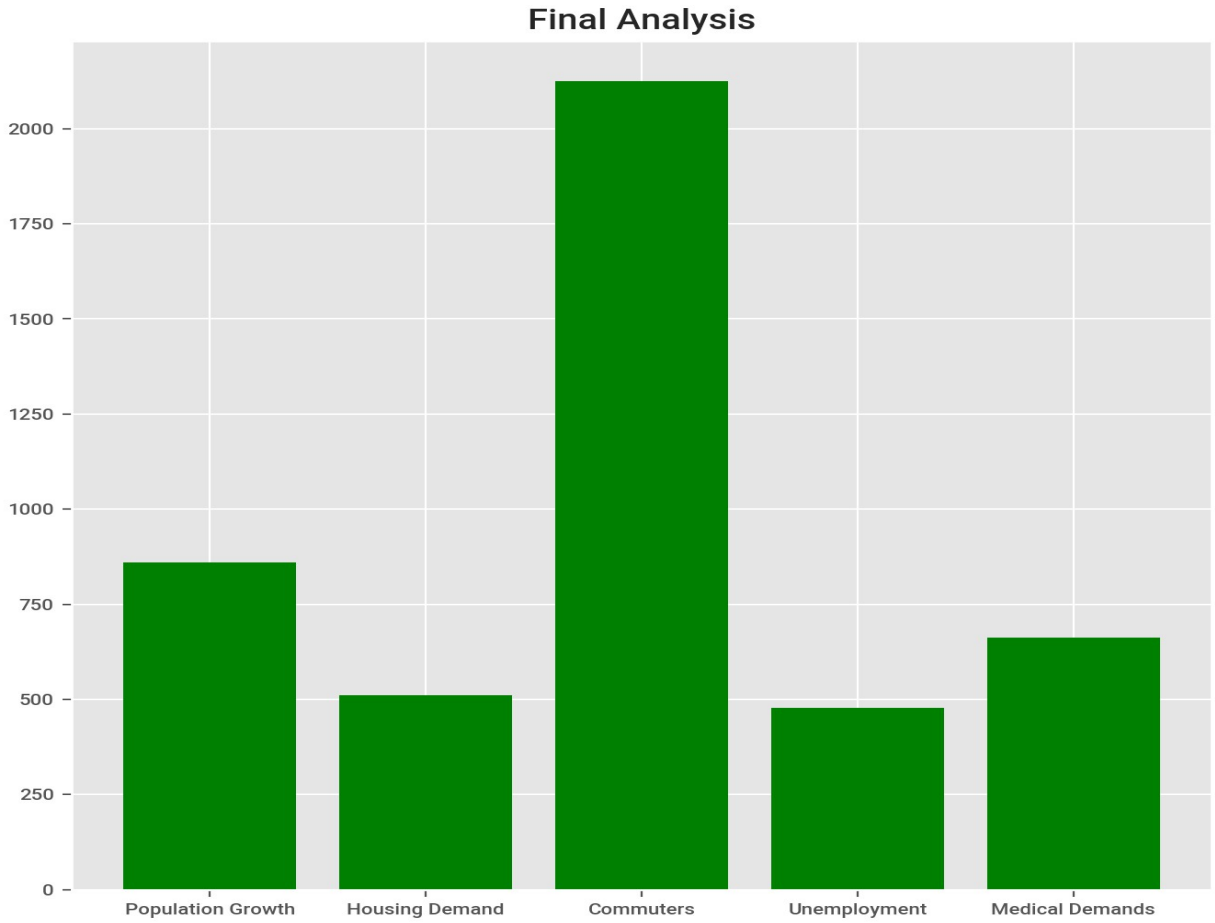
```
print("\033[1m\"Final Analysis for Demand for Housing\"\033[0m","\n")

print("The total number of people living in houses more than the street average occupancy are : %d \n"% abs(neg))
print("The total number of residential space available(buffer) compared to the streets house average occupancy are : %d \n "
#Lets assume that the residential space available is the potential buffer that can be used for accomodating the people satur
print("Requirement of housing space is : %d \n"%abs((pos+neg)))
print("Requirement of housing space for 100K people frame is : %d \n" % abs(housing100k))
print("\033[1mDo we need more housing ? : Yes \033[0m")
```

```
"Final Analysis for Demand for Housing"

The total number of people living in houses more than the street average occupancy are : 1749

The total number of residential space available(buffer) compared to the streets house average occupancy are : 1238

Requirement of housing space is : 511

Requirement of housing space for 100K people frame is : 6170

Do we need more housing ? : Yes
```

Recommendations

Based on the above presented data analysis of the town population it is recommended to build a Train station since there are potentially a lot of commuters in the town and building the train station could take the pressure off the roads . Also builtinf more low density housing can ease off the load of the lodgers coming to the town , presently the town is growing so there might be more people coming in this town for work or settlement.

Given that the population of the town is increasing and there is a considerable growth since a significant number of lodgers and visitors are flowing into the town , it is suggested to consider in investing in General infrastructure as this will benefit more people generally. However the aging population will also increase and will require more care so it is also suggested to invest a fraction into old age care.



Bibliography

Eurostat (2017) *Marriage and Divorce Statistics*
Available online: https://ec.europa.eu/eurostat/statistics-explained/index.php/Marriage_and_divorce_statistics

Eurostat (2020) *EU crude divorce rate on the rise*
Available online: https://ec.europa.eu/eurostat/web/products-eurostat-news/-/ddn-20200710-1