

# United Kingdom

## 1881 Census Project Report

### Abstract

The report focuses on the analysis of the census of a small town, carried out in the United Kingdom in 1881 to ask important developmental questions, prepare, process, analyze, and make informed data-driven decisions for future economic, social and infrastructural planning of the community.

### Introduction

The purpose of the census is to compare different demographics of the nation, and provide the government with accurate statistics to enable better planning.

Some common factors such as household population, number of commuters, birth rate, death rate, Marital Status, religion, the number of seniors, population growth was assessed. In recent times, net migration has directly contributed to more than half of population growth (Cangiano, 2019). It also analyzes the current state of health of the nation by its infirmity rate.

This report highlights employment and unemployment trends by assessing the working age population, their occupation, type of occupation and how it affects the town's economic output. It also examined and evaluate the current state of the population to identify investment opportunities for growth, deficits and how to deal with existential challenges

.

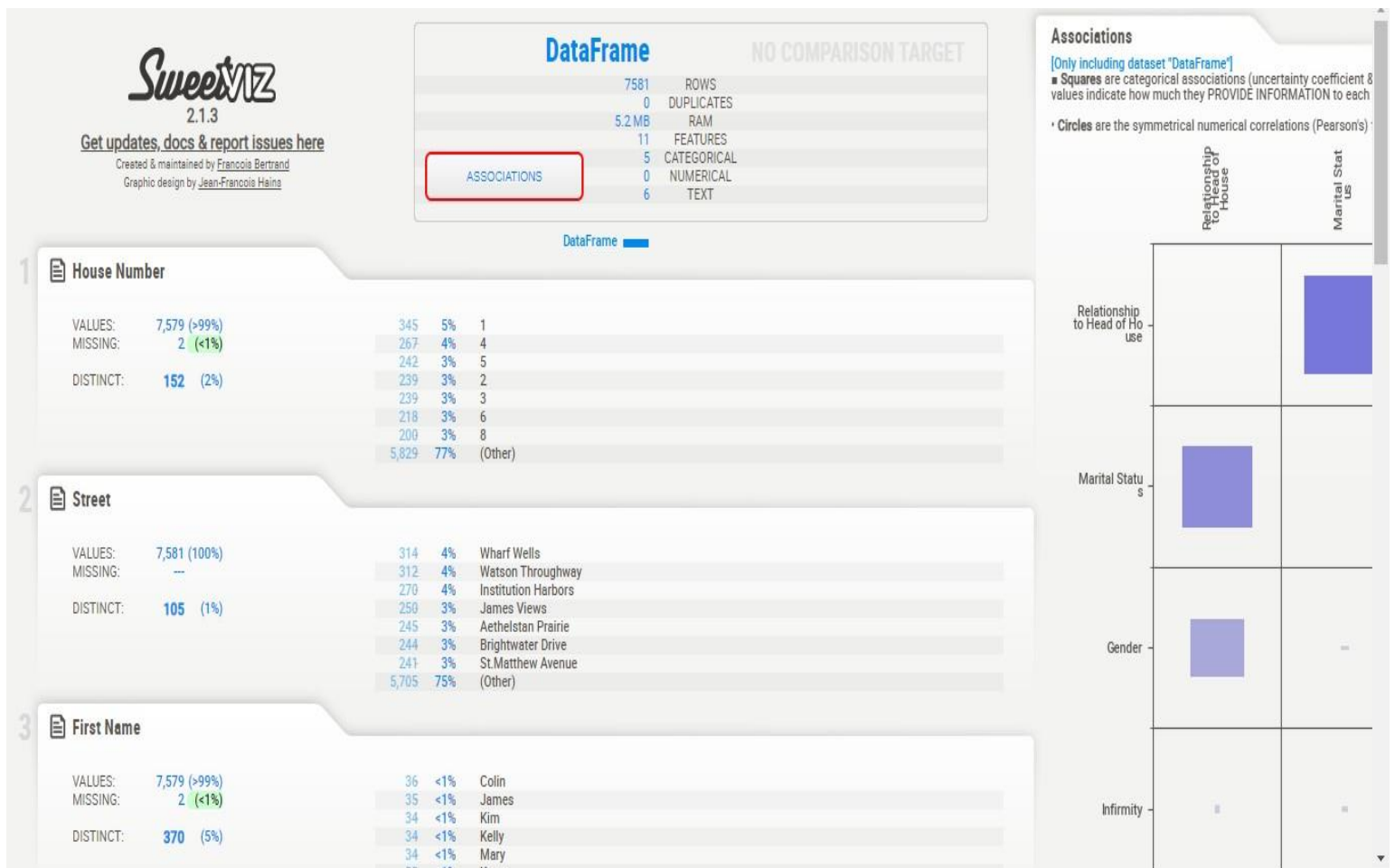
### Data Cleaning

Census data was cleaned to remove duplicate or irrelevant observations, fix structural errors, filter and replace unwanted outliers, handle missing values, manage nan values, detect inconsistencies in order to improve the overall quality of data (Rahm, Hai Do, 2000).

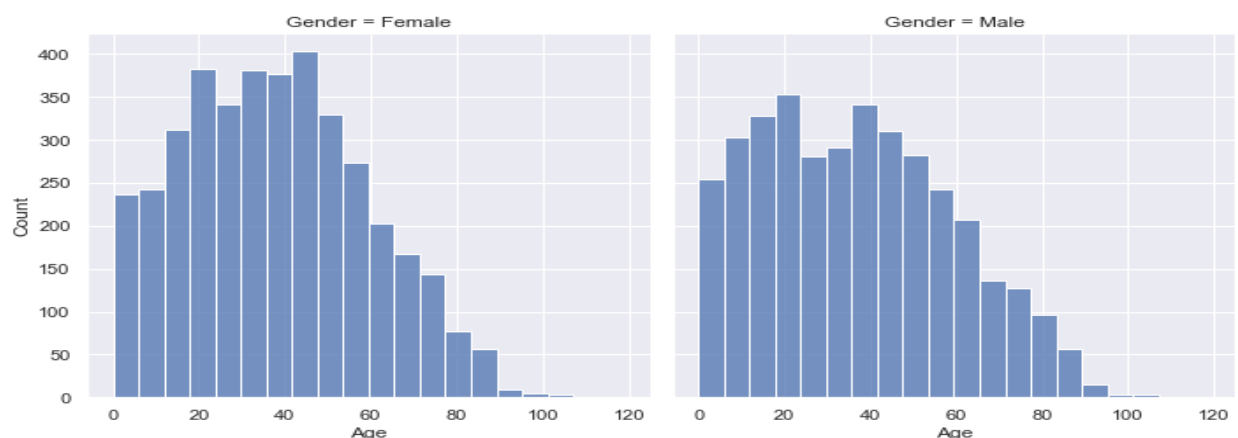
Data was manipulated on Jupyter Notebook using python programming modules such as pandas, seaborn, and various other libraries to accurately pre-process data.

The data was analyzed at its raw state using an exploratory data analysis module called sweetviz to view and note some obvious problematic areas such as the missing values, columns with blank spaces, and note unique values. It also shows some critical aspect that will be useful further in my analysis. Aspect such as "Association" of different field to one another.

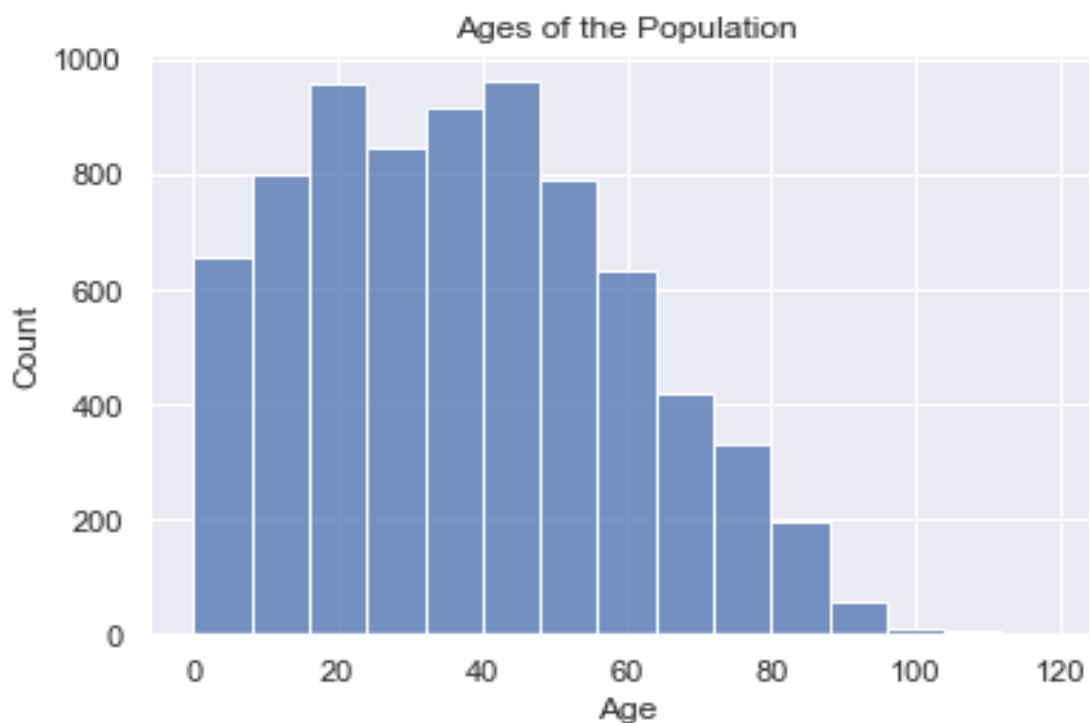
Inspecting the missing values and data types of each column. Religion, 1717, had the most null values, followed by marital status, 1677, and then we have two each in surname and house number respectively. Only one field had a type that needed to be changed and that was Age which was an object.



**Age:** This had strings, a negative age and floating-point numbers. First, the word numbers ('Four' and 'One') and the negative number were identified and replaced with their numeric values and zero. Next, investigation showed that there were no empty strings in age. Rounded all ages from floating point numbers to a single digit, as age cannot be a float. The field was then cast from object to float and then to the rightful data type integer. The oldest human in modern era was 122years 164 days before she died, so the maximum age in my dataset which was 131 was replaced with 107; the age of the individual's



partner as suggested by the Surname, house number, marital status and relationship to the head of house column.



**Religion:** The Charity Commission for England and Wales has ruled that Jedi is not a religion (The Charity Commission, 2016). I replaced Jedi and Private with unknown in the field, replaced adults aged over 17 with NaN religion with “None” and replaced children under 17 years old with NaN religion with Undeclared in the data series.

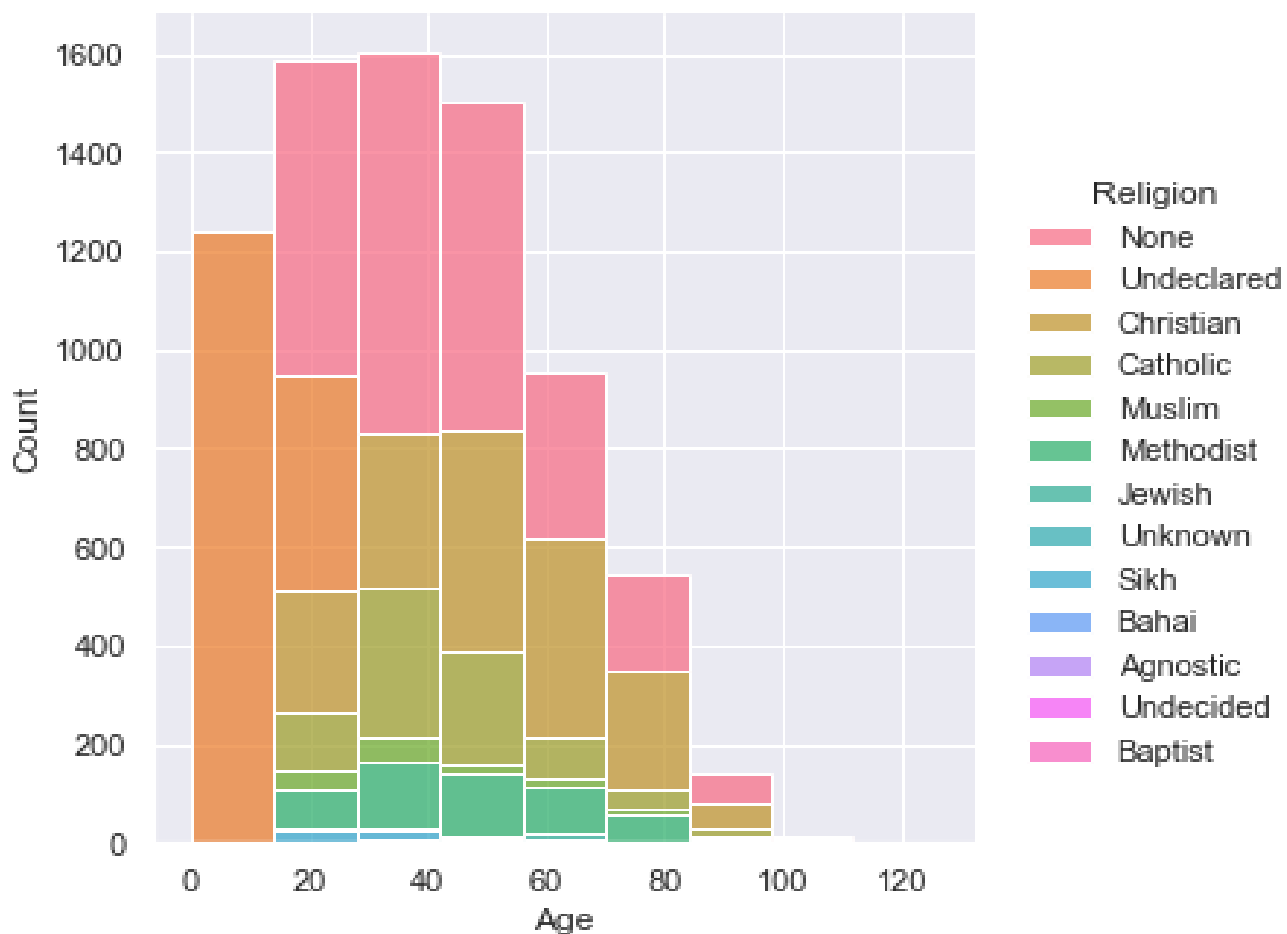
Religion has three blank spaces, Line 2432 in the census data suggests he deliberately left it blank, Line 7000, looking at the spouse (7001), we can carefully assume they chose **None**.

The individual in line 7034 seem to live in Christian household but that does not suggest where she stands as she is an adult (age: 58). So, I went with Undeclared.

Computing the mean age of each religion as declared showed that the choice of Christianity average age is 50; the highest amidst the lot alongside Bahai. Highest median age falls on Christianity as well at 52. “None” is the mode of the religion field and 19-year-old are the most frequent with the choice of “None”.

The Interquartile range is 34.

	count	mean	std	min	25%	50%	75%	max
Religion								
Agnostic	5.0	25.000000	3.674235	22.0	22.00	25.0	25.00	31.0
Bahai	3.0	50.666667	17.559423	34.0	41.50	49.0	59.00	69.0
Baptist	2.0	40.000000	1.414214	39.0	39.50	40.0	40.50	41.0
Catholic	794.0	42.528967	15.489973	18.0	32.00	40.5	49.75	103.0
Christian	1713.0	50.824869	18.684222	18.0	36.00	52.0	64.00	119.0
Jewish	35.0	45.800000	20.876105	21.0	27.00	38.0	61.50	88.0
Methodist	506.0	47.043478	17.838830	18.0	34.00	46.0	60.00	104.0
Muslim	127.0	38.669291	16.605151	18.0	26.00	33.0	50.00	89.0
None	2664.0	42.474850	17.691332	18.0	28.00	40.0	53.00	106.0
Sikh	49.0	36.632653	17.713806	18.0	24.00	29.0	48.00	88.0
Undecided	2.0	32.500000	3.535534	30.0	31.25	32.5	33.75	35.0
Undeclared	1679.0	9.216796	5.501928	0.0	5.00	10.0	14.00	75.0
Unknown	2.0	49.000000	11.313708	41.0	45.00	49.0	53.00	57.0



**Infirmary:** Five blank spaces in this column was replaced with “None”. Investigating their respective household, it is safe to assume they are of good medical condition as all members of these households chose none. So blank spaces were replaced with None.

**Gender:** The Gender field had some inconsistencies in the submissions such as Female and Male written in different variations like F, Fe-male, M. All were replaced with a consistent version which are Female and Male respectively. Two blank spaces were located in the column and were replaced consistent with their respective gender as signaled by their first names.

**Occupation:** Occupation had no data cleaning need but it was vital to the overall outcome our most of our analysis. See below.

**Marital Status:** There was an inconsistency with the status divorced written as Di-vorced, this was replaced with "Divorced". We have a large volume of Null values; most are under 17. All NaN values 13 years and above was replaced with the mode of the Marital Status column which is "Single. This choice was made as most teenagers, in most household from 13 years of age are allowed to date. One adult with NaN marital status was also replaced with "Single. One blank entry aged 13 was replaced with single. The Null entries (1143) left are children, under 13 and are not considered/assumed to have any status.

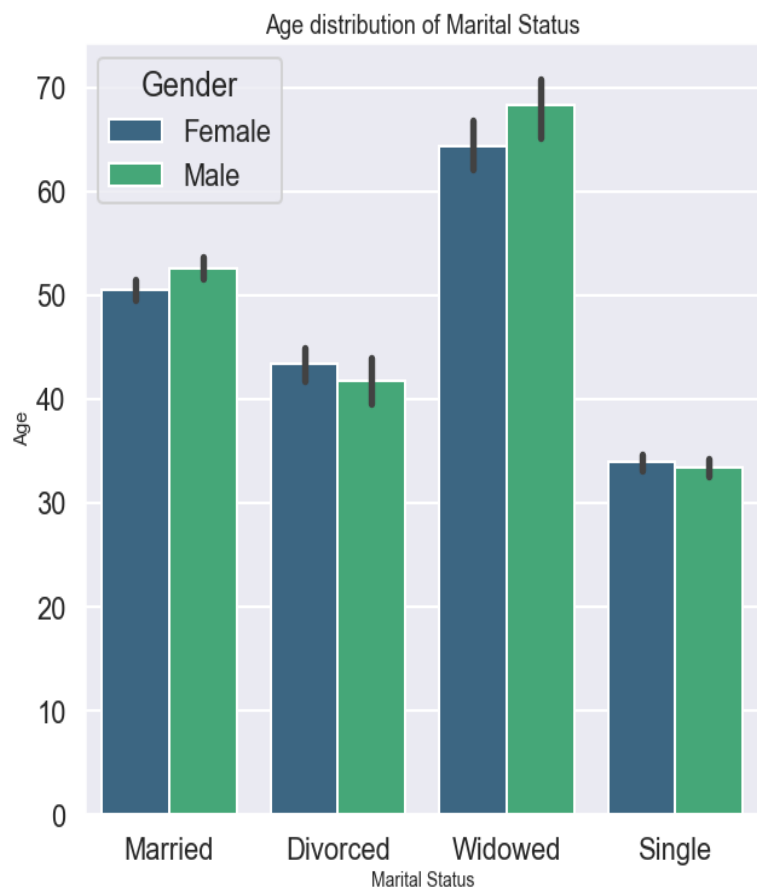
	count	mean	std	min	25%	50%	75%	max
<b>Marital Status</b>								
Divorced	751.0	42.727031	17.399503	18.0	29.0	40.0	53.0	94.0
Married	2182.0	51.587076	17.424323	18.0	39.0	50.0	64.0	119.0
Single	3130.0	33.677955	15.453240	13.0	20.0	31.0	45.0	104.0
Widowed	375.0	65.922667	18.023808	18.0	60.5	71.0	77.0	105.0

**Relationship to Head of House:** From investigation, no individuals under 17years old recorded was head, husband, partner, wife, son-in-law.

**Surname:** Row 5058 in Surname had a missing data, it was replaced with the mode of the column, "James", consistent with the gender of the individual.

**First Name:** Here, there are two incomplete names that are N and P respectively. These were replaced with the mode of the first name - both are male as seen in the Gender column. Two Null values were replaced with the mode of their respective gender ("Kim", "Colin") with respect to the first name field.

**House Number:** Word numbers were replaced with their numeric characters. There were two NaN values. Row 7366, the data suggests he is part of the Murphy family at 13 chester Lane. Row 7445, the data suggests she is part of the Bibi family at 36 chester Lane. They were replaced with 13 and 36 respectively.



## Analysis

**Train station.** There are potentially a lot of commuters in the town and building a train station could take pressure off the roads.

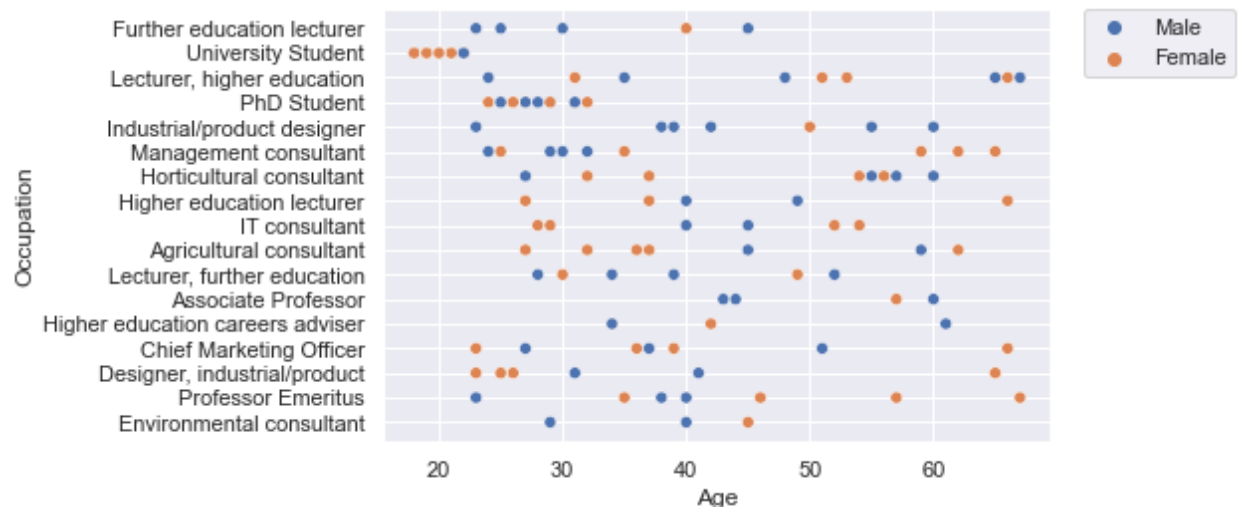
The town' as shown by its numbers and somewhat lack of some facilities such as a university that there will be commuters who frequently travel for work.

To begin the analysis, occupation field was filtered for commuters for three different types of people:

- University and PhD students.
- University lecturers and other University staffers.
- Consultants, Marketers, and technical personnel in Information Technology.

Lecturers, university students and some other fraction of occupation whom because of the dynamics of their jobs might need to travel for work often. Examples are consultants and marketers. See table extract below:

Occupation	count	mean	std	min	25%	50%	75%	max	count	mean	std	min	25%	50%	75%	max
Agricultural consultant	7.0	19.714286	15.184579	6.0	11.00	11.0	24.50	50.0	7.0	42.571429	13.427052	27.0	34.00	37.0	52.00	62.0
Associate Professor	4.0	11.250000	14.930394	1.0	1.75	5.5	15.00	33.0	4.0	51.000000	8.755950	43.0	43.75	50.5	57.75	60.0
Chief Marketing Officer	7.0	24.857143	24.113817	1.0	2.00	20.0	49.00	51.0	7.0	39.857143	14.610824	23.0	31.50	37.0	45.00	66.0
Designer, industrial/product	7.0	20.285714	22.997930	1.0	3.50	5.0	36.00	57.0	7.0	33.714286	15.085471	23.0	25.00	26.0	36.00	65.0
Environmental consultant	3.0	32.333333	34.312291	1.0	14.00	27.0	48.00	69.0	3.0	38.000000	8.185353	29.0	34.50	40.0	42.50	45.0
Further education lecturer	5.0	30.400000	29.720363	1.0	6.00	24.0	50.00	71.0	5.0	32.600000	9.555103	23.0	25.00	30.0	40.00	45.0
Higher education careers adviser	4.0	35.250000	22.381168	4.0	29.50	40.0	45.75	57.0	4.0	42.750000	12.737739	34.0	34.00	38.0	46.75	61.0
Higher education lecturer	5.0	18.800000	14.669697	4.0	9.00	15.0	25.00	41.0	5.0	43.800000	14.686729	27.0	37.00	40.0	49.00	66.0
Horticultural consultant	8.0	21.500000	16.877711	3.0	9.75	15.0	35.00	46.0	8.0	47.250000	13.024702	27.0	35.75	54.5	56.25	60.0



The total according to this commuter's travel analysis is 600 people travelling out of town to work, school, every day.

This is 7.91% of the population which is not really a substantial number to arouse a need to invest limited resources into train station but in addition to this, a significant number of people are employed (52.7%), plus college students, secondary school student and normal everyday business owners who commute to school every day within the town, makes it almost 65-70 percent of commuting population.

This shows that there is an urgent need for investment in public transportation system to alleviate the pressure off the roads and reduce stress of the people. See fig 8: Employed Chart.

**Seniors:** The numbers of seniors in a society are one of future economic indicators. The figure of seniors in our census data stands at 800, this is 10.55% of the total population. This is not where the concern is; in modern times local governments where more than 21.6% of the population are aged 65 years and over), (ONS, 2018). According to our Age pyramid, it clearly shows that people of age 30-64 in 30years will be seniors, that is a huge 47.25% of the population in and towards retirement. This shows that there will be high demand for financing of **old age health care, higher pension funding**.

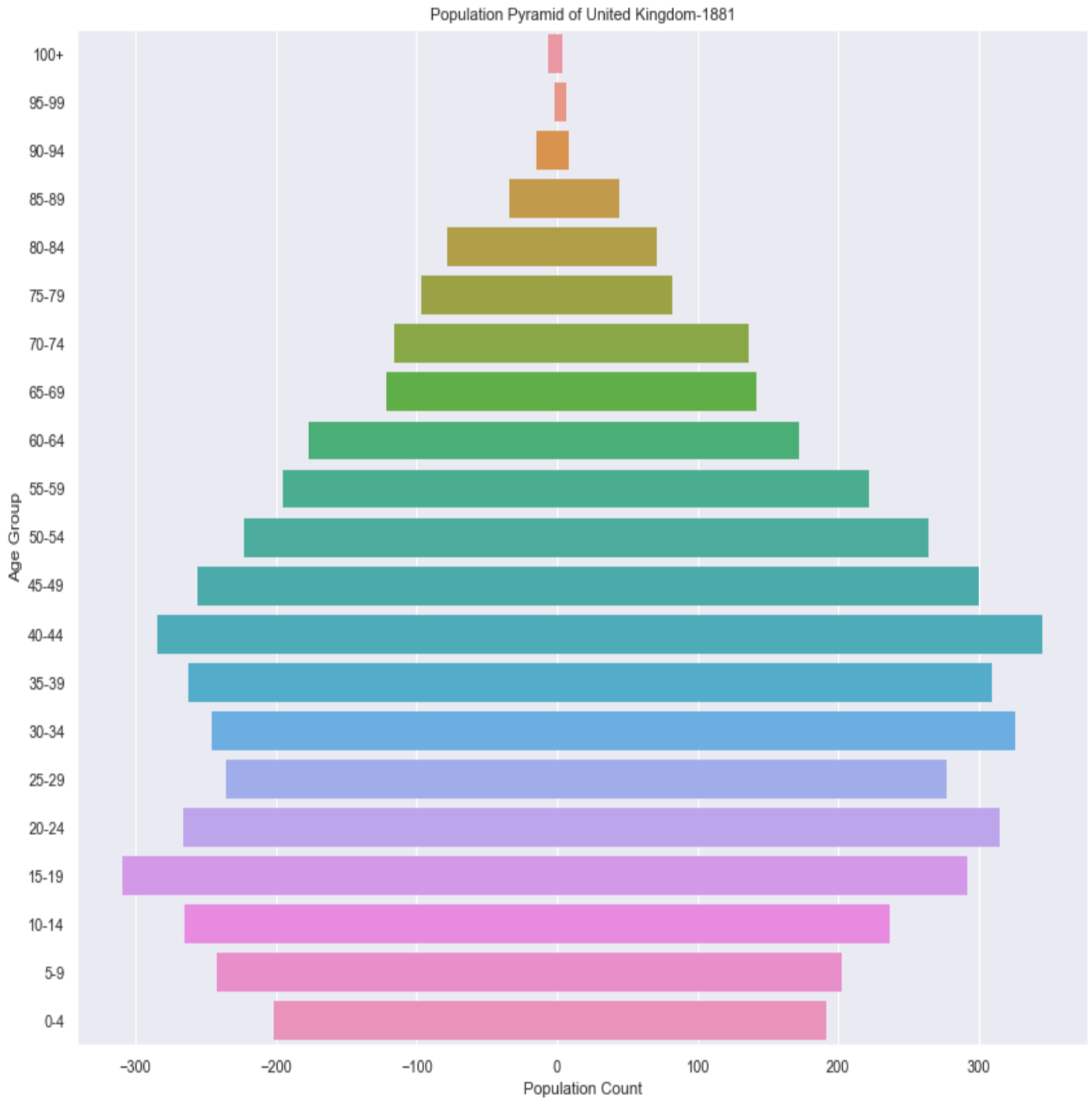


fig. Age Pyramid

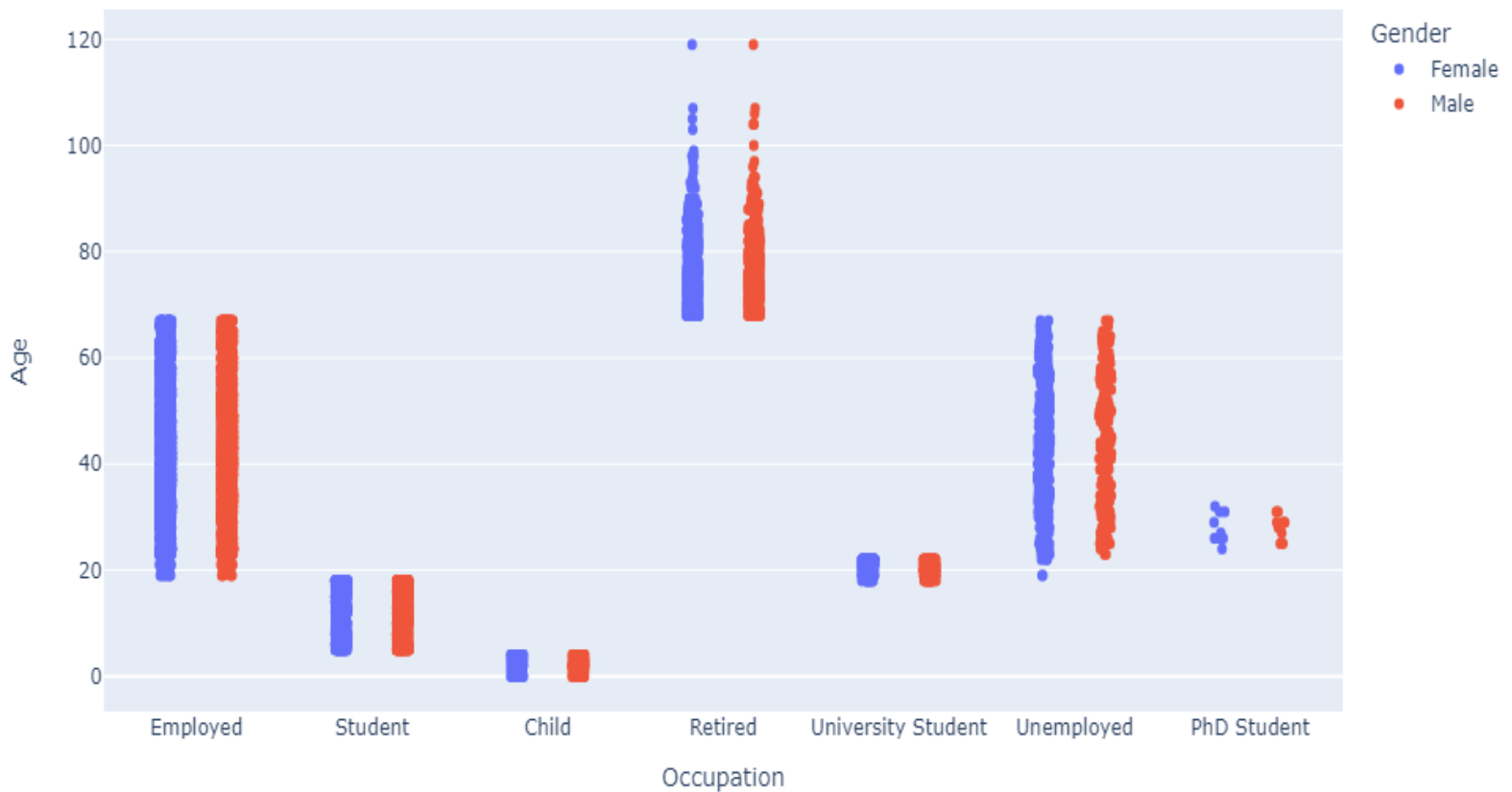
### Unemployment Trend:

The headline unemployment rate is calculated by dividing the unemployment level for those aged 16 and over by the total number of economically active people aged 16 and over. Economically active is defined



as those in employment plus those who are unemployed. (ONS, 2020). Several categorizations were made from the occupation table:

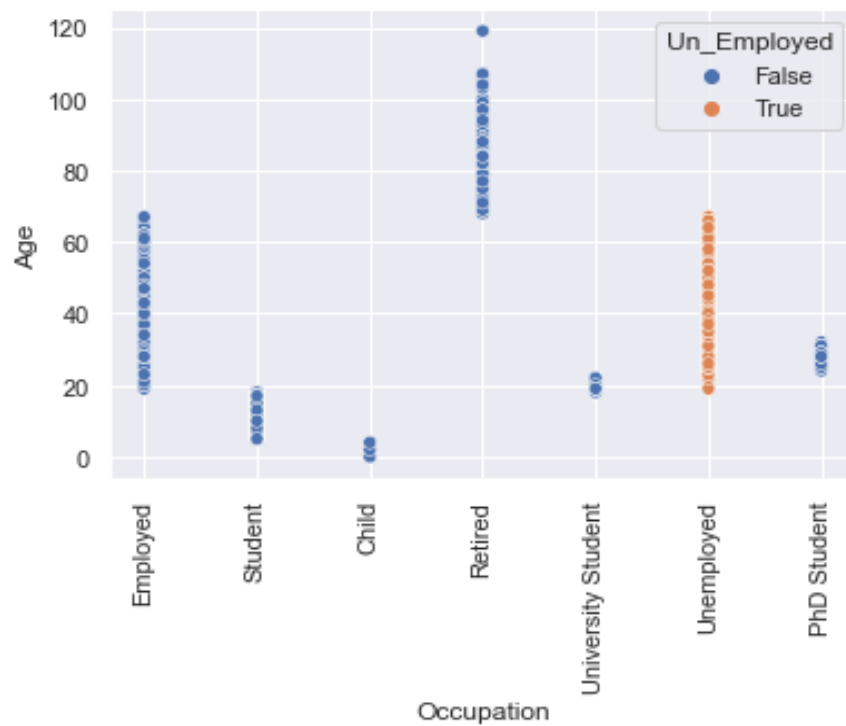
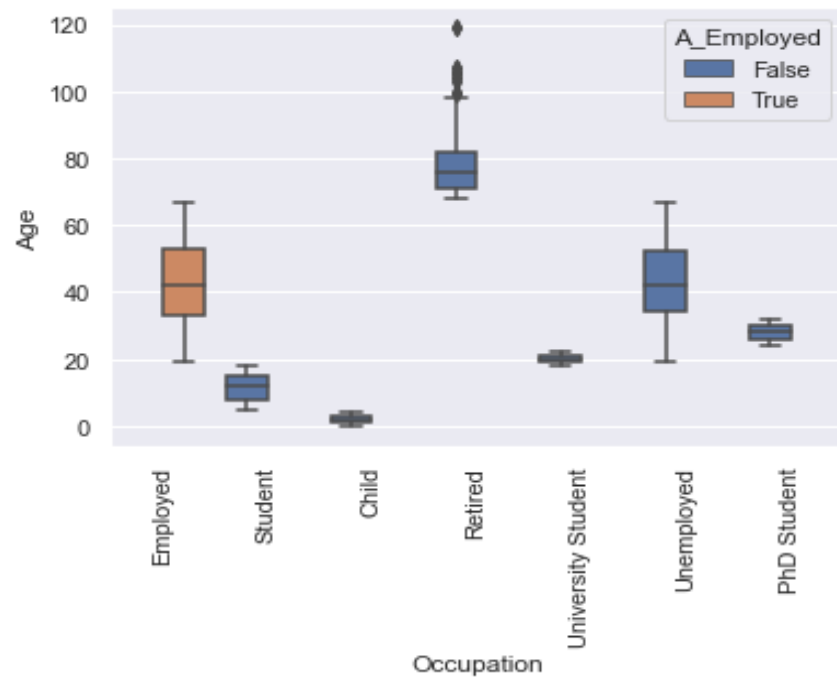
The chart below shows the overall occupation chart conditioned by gender. See below:



First, our data shows that aged 68 and above are retired. A “Retired” category was made (see Seniors). We have 800 retirees in our data.

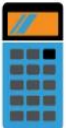
- Another category was filtered from the population called “Employed”. These are people currently gainfully employed. 3997 is actively employed from our data.
- The third category is the Unemployed. 503 actively unemployed from our data

Economically active people were calculated by adding the number of those in employment plus those who are unemployed. Total is 4500.





	Employed	Unemployed
0	Employed	3997
1	Unemployed	503

To calculate the unemployment rate:



**Unemployment  
Rate Formula**

$$= \frac{\text{No. of Unemployed Persons}}{\text{No. of Employed Persons} + \text{No. of Unemployed Persons}}$$

The unemployment rate is 11.8%.

### High-density housing.

High-density housing should not be built on the unoccupied land. Projection shows that there will be a need to build this in the next 5 years. Here's why:

Immigration levels – conditioned by people that are likely to stay because of employment.

The emigration levels, looking at the number of divorcees and their likelihood to stay back in town or 'not' because of unemployment.

Occupancy rate is also a contributor to this decision. The House number and street was grouped to find the degree of people living in each household of the population. See table below:

House Number		Street	Occupancy Degree
0	1	Adams Ranch	5
1	1	Adams Ways	1
2	1	Aethelstan Prairie	2
3	1	Albert Lane	6
4	1	Albion Springs	1
...	...	...	...
2873	146	James Views	1
2874	147	James Views	2
2875	148	James Views	2
2876	149	James Views	1
2877	150	James Views	1

The average(mode) of people living in a household is two (2).

Checking this against the number of people currently living in each household, the density stands at 1207 to 1671. An average of four people living in a household meant for two.

The decision to not built high density houses is because the town will not have a housing crisis in the short term because of low migration levels and birth rate but should experience deficit in 2-5 years. My recommendation is that the local government authority should look to

make provisions in the next calendar year budget to build high density houses as there is a slight increase in population.

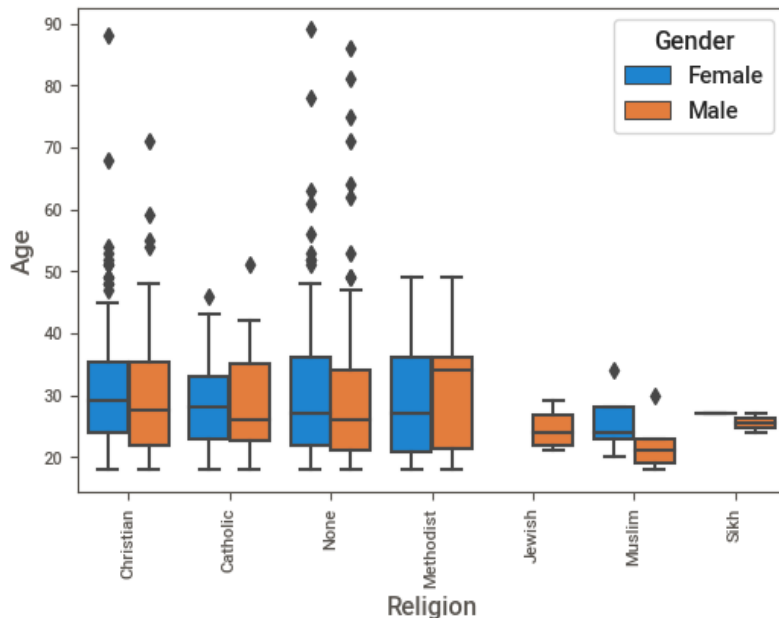
House Number		Street	Occupancy Degree	Actual_Occupants
0	1	Adams Ranch	5	-3
1	1	Adams Ways	1	1
2	1	Aethelstan Prairie	2	0
3	1	Albert Lane	6	-4
4	1	Albion Springs	1	1
...	...	...	...	...
2873	146	James Views	1	1
2874	147	James Views	2	0
2875	148	James Views	2	0
2876	149	James Views	1	1
2877	150	James Views	1	1

#### Religious building:

There is already one place of worship for Catholics in the town. This data presents that there is no immediate need for more religious buildings. The statistics shows that there is a growing population of **none religious**. It also shows that many who chose not to declare their religious affiliation, come mostly

from a Christian, Catholic background. Religious affiliations was checked especially people practicing the same religion as the parents and how it affects the growth of religion

A number of people over 17 who lives in the same household with their parents or guardian share same religion. Although 870 people fall within this category but in the immediate short term, there will not be a need for more religious buildings.



Other religion is showing signs of natural growth as well but not significant enough for a new building.

	House Number								Age							
	count	mean	std	min	25%	50%	75%	max	count	mean	std	min	25%	50%	75%	max
Infirmity																
Blind	5.0	19.800000	21.335417	1.0	10.00	12.0	20.0	56.0	5.0	59.400000	18.609138	40.0	43.00	59.0	71.00	84.0
Deaf	6.0	19.333333	7.659417	10.0	13.75	18.5	25.5	29.0	6.0	40.000000	30.033315	4.0	17.75	37.0	63.75	78.0
Disabled	2.0	14.000000	15.556349	3.0	8.50	14.0	19.5	25.0	2.0	17.500000	14.849242	7.0	12.25	17.5	22.75	28.0
Mental Disability	5.0	21.400000	19.781304	3.0	5.00	21.0	26.0	52.0	5.0	49.200000	34.346781	18.0	22.00	37.0	71.00	98.0
None	7545.0	26.822001	25.269339	1.0	8.00	20.0	37.0	150.0	7545.0	37.200795	22.016873	0.0	19.00	36.0	53.00	119.0
Physical Disability	13.0	32.153846	31.955297	2.0	14.00	25.0	37.0	124.0	13.0	35.230769	24.056024	1.0	20.00	34.0	45.00	98.0
Unknown Infection	5.0	17.400000	12.992305	6.0	8.00	10.0	30.0	33.0	5.0	27.200000	22.698018	7.0	11.00	15.0	46.00	57.0

### Emergency Medical Building:

There is also no need for minor injury center at this time as the rate of infirmity is very low at 0.2%. The birth rate also only has a natural increase, so, there will be no immediate in the short term. The table above shows that 99.52% of the population recorded no infirmity.

## Divorce and Marriage rate

Divorce rate is the number of divorces divided by the population of married people. From our data,

- Number of divorcees: 751
- Number of Married people is: 2182

Divorce rate per 1000 people over Married, gave us 344.18 - 34.4% of population. If this keeps up its pace, The town will need an intervention for housing projects as more people will live alone than together.

## Birth Rate, Death rate and Difference

The birth rate (BR) is equal to the number of live births in a year divided by the total population of females within child bearing age (25,34), multiplied by 1,000.

This is the statistics:

- Live births in census data: 70
- Women between 25-29 years with new born: 277
- Birth rate of women between 25-29 years old: 252.7075812274368
- Four-year-old in census data: 86
- Women between 30-34 years with 4-year-olds: 325
- Birth rate of women between 30-34 years old: 265
- Birthrate over a 5-year period for the two-age range is: 517.7075812274368
- Birth rate is for one calendar year: 60.265681112931915 children

The general birth rate as shown above is 60 per 1000 of the population. Calculating death rate, 36 died in the calendar year.

Natural increase in population is calculated by subtracting death rate per 1000 of population from birth rate per 1000 of population and then divide by 10. This can also be calculated by subtracting deathrate per 1000 from birth rate per 1000 divide by 1000 multiplied by 100 (%).

$$\frac{\text{\#births in 1 year} - \text{\# deaths in 1 year}}{\text{\# thousand total population}} = \text{Rate of Natural Increase}$$

Natural increase in population census is: 2.4%

## Is the Population Expanding?

The three main causes of population change.

- Births - usually measured using the birth rate (number of live births per 1,000 of the population per year).
- Deaths - usually measured using the death rate (number of deaths per 1,000 of the population per year).
- Migration - the movement of people in and out of an area.

## THE DEMOGRAPHIC EQUATION

The basic demographic equation is:

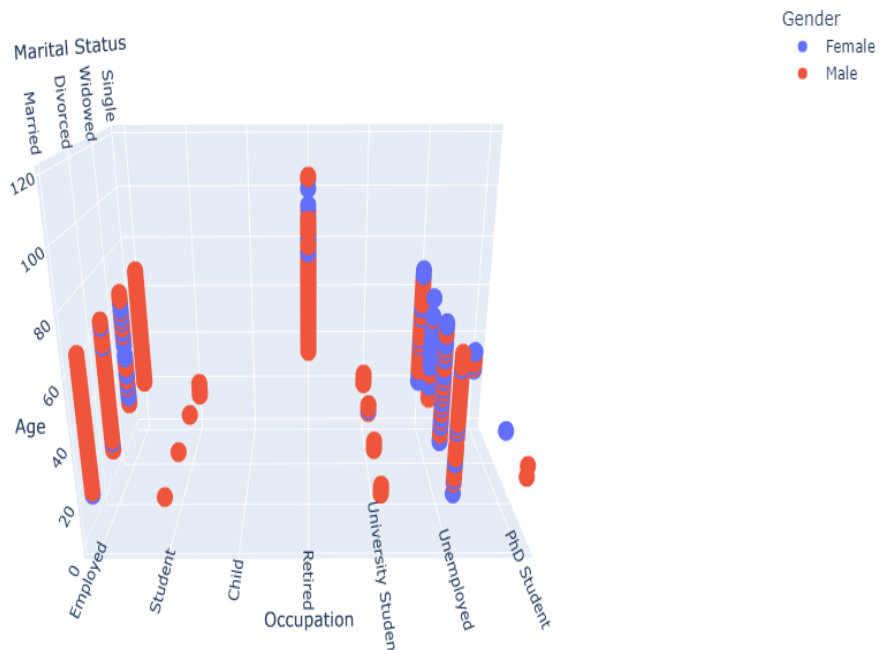
$$\text{Migration in} - \text{migration out} + \text{births} - \text{deaths} = \text{total population change}$$

(Scharks, 2013)

$$\begin{aligned}\text{Population Change} &= 36.4 - 17 + 60.3 - 36.4 \\ \text{Population Change} &= 43.3\end{aligned}$$

## Life expectancy:

Life expectancy is the number of years a person can expect to live. By using the age range which people die in our census as a cohort life expectancy, we can deduce two things, people grow old in this town, especially the married and gainfully employed until retirement.



## Reference

Impact of migration on population, Cangiano, 2019, Source: [Impact of Migration on United Kingdom Growth](#)

Data cleaning, ([Rahm](#), 2000) - [Data cleaning Research Paper](#)

Jedi is not a religion, (Charity Commission for England and Wales, 2016) – [Charity Commission for England and Wales The Temple of The Jedi Order Application for Registration Decision of the Commission 2016](#)

Retirement age, (uk.gov, 2021) – [Retirement Pension Age](#)

Aged people, (ONS, 2018) - [Office of National Statistics, United Kingdom](#)

Unemployment rate Image, (Thakur, 2018) - <https://www.educba.com/unemployment-rate-formula/>

Population Growth, (BBC, 2021) – [Cause of Population Change](#)

Population Change (Scharks, 2013) – [Demographic Equation of Population Change](#)

Immigration and Emigration rates – (Palumbo, 2014) – [Calculating Migration Rates](#)