

Project_Report-5.docx

by RAUNAK MASIH

Submission date: 23-Apr-2024 08:49PM (UTC+0100)

Submission ID: 230536487

File name: Project_Report.docx (273.09K)

Word count: 2496

Character count: 12832

Fundamentals of Data Science
Project Report on Census Data

1. Introduction

This project aims to analyze census data from a small town and use the results to guide decisions on what should be built on an undeveloped property that the local government wants to develop and how much money should be allocated to community services. Following a thorough data cleaning procedure that included finding and addressing shortcomings in the dataset, analysis was performed on the census data.

The specifics and procedures used throughout the data cleaning process are given in the report's next section. Subsequently, the analysis's output is given, offering statistical insights pertinent to the process of making decisions. Finally, suggestions for investments are given, and these are predicated on the analysis's findings.

2. Data Cleaning

Data must first be verified for errors and inconsistencies and these must be corrected using a variety of techniques before it can be properly analyzed. This procedure is known as data cleaning, and for this project, a thorough review was done on each and every column in the dataset. The Jupyter notebooks implemented in Python environment were used for the analysis. **Figure 1** shows the summary of the Dataset and the eleven columns in the dataset comprised variables that provide comprehensive details on the town's residents. Majorly, four essential steps were established to help find and correct problems in the dataset across all columns. These are as follows:

- Assessing the column's unique values.
- Verify the column's datatype.
- Determine whether a column contains null values.
- Verify whether a column's string values are empty.

Figure 1 : Dataset Summary

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10118 entries, 0 to 10117
Data columns (total 11 columns):
#   Column              Non-Null Count  Dtype
---  -
0   House Number        10118 non-null  int64
1   Street              10118 non-null  object
2   First Name          10118 non-null  object
3   Surname             10118 non-null  object
4   Age                 10118 non-null  float64
5   Relationship to Head of House  9356 non-null   object
6   Marital Status      7634 non-null   object
7   Gender              10118 non-null  object
8   Occupation          10118 non-null  object
9   Infirmary           84 non-null     object
10  Religion            4123 non-null   object
dtypes: float64(1), int64(1), object(9)
memory usage: 869.6+ KB
```

The following is representation of the data cleaning procedure:

Column Age: The values in the column were converted from floats, the incorrect datatype, to integers.

Column Gender: The column included empty string values, which were resolved by comparing the value in the relationship to head of house column to the same House Number and Street and utilizing that value to determine the value in this column.

Column Relationship to Head of House: This column contained empty string, missing values as well as spelling error wherein "Neice" was spelled incorrectly; it was corrected to "Niece." This missing value column is filled by checking if the 'Relationship to Head of House' for the same 'House Number' and 'Street' is already filled, then for missing values assign 'Child', 'Spouse', or 'Lodger/Visitor' based on age, marital status, and gender. For example, if age is less than 18, assign child but also checks for gender mismatch. For example, if daughter in existing value and gender is male or son in existing value or gender is female, assign 'Child' instead. Empty string filled by value 'Unknown'.

Column House Number: This column was error-free.

Column Street: This column contained no mistakes.

Column First Name: This column was error-free.

Column Surname: The row with the blank string was corrected by associating the specified surname with the empty string and verifying that the house number and street were the same.

Column Infirmary: This column contained both null and empty string values; to fill in the NaN values, it was altered to Unknown. Since Unknown currently makes up 99% of the data in the column, it is also used to fill in the empty strings.

Column Occupation: This column was error-free.

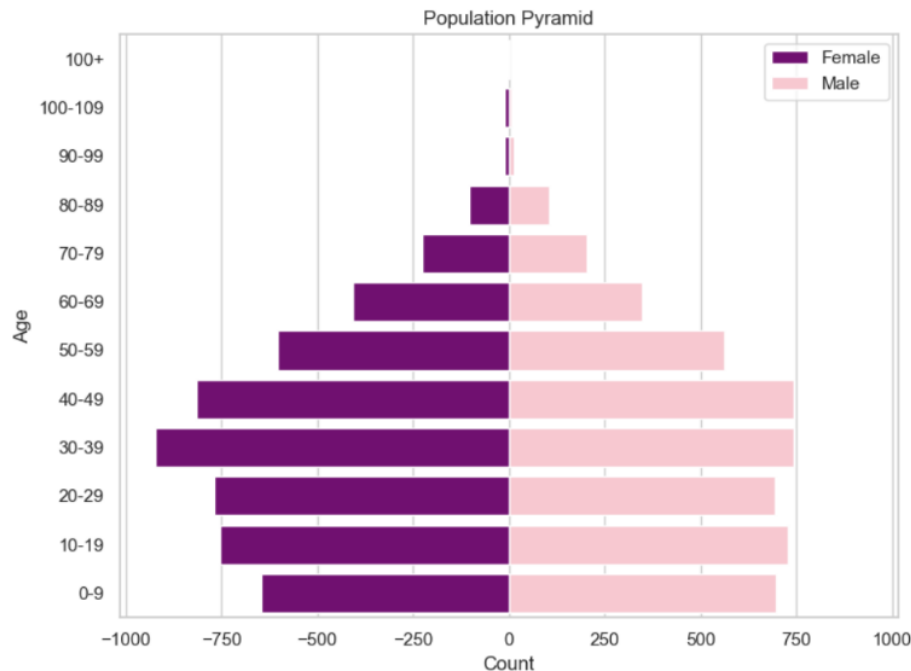
Column Religion: This column had a mix of errors including empty string values and null values, these null values are filled by grouping the same House Number and street. If there's only one unique religion within same House Number and Street, return that religion; otherwise, returning 'Undecided'. Additionally, empty strings are filled by value 'Undecided' also 'Nope' is changed to 'Undecided'.

Column Marital Status: This column is filled by checking if the 'Marital Status' for the same 'House Number' and 'Street' is already filled. Furthermore, we check if it contradicts the 'Relationship to Head of House'. For example, if the existing value in the 'Marital status' is Divorced and Relationship to head of house is either 'Wife' or 'Husband' or either the existing value in marital status is Married and relationship to head of house is neither 'Wife' nor 'Husband', assign 'Unknown'. However, if there is no contradiction, fill with the existing value. Additionally, there is one condition to check whether if age is less than 18, assign it to be 'Single'. Empty strings are filled by value 'Unknown'.

3. Demographics Age Pyramid

The population of the town is 10,117. Of the total population, 4,848 (47.91%) are men and 5,269 (52.08%) are women. As seen in Figure 2, the age range that is highest among all is 30 to 39. The town's median age is 34 years old overall.

Figure 2: Population Pyramid



3.1. Birth Rate and Death Rate

It is believed that there are 110 children in the town who are zero years old, which represents the number of live births. With this in mind, 29.6 births per hundreds is the anticipated crude birth rate for women in the town between the ages of 25 and 29. For the town, the approximate crude death rate is 8.85 deaths per hundreds. The birth and death rates in the United Kingdom are derived from National population projections, fertility assumptions: 2020-based interim (Office for National Statistics, 2020) shows that total fertility rate (TFR) in the UK has been falling since 2012, when it was 1.92, and is estimated to be 1.56 in 2020.

3.2. Migration

The released data on net migration to rural areas in the United Kingdom (Social Mobility Commission, 2020) highlights the socio-economic origins of migrants, pointing out that people from wealthy homes are more likely to relocate for employment or education. A net internal migration outward of 149,100 persons occurred in predominantly urban areas

between 2019 and 2020. Additionally⁴ according to the data estimated there was an average net migration of 6.05 per thousand between 2011 and 2020 (Department for Environment Food & Rural Affairs, 2021).

¹

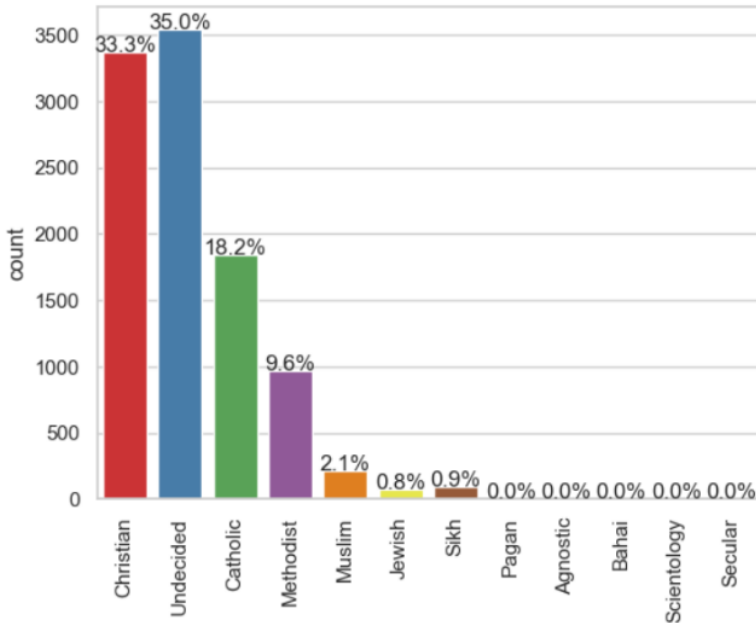
3.3. Estimated Population Growth Rate

The town estimated population growth rate, which takes into account both the present population growth rate and the growth rate, is 1.0%. Data for UK show that between 2011 and 2019, the population of the UK increased in both rural (5.2%) and urban (6.2%) areas, with the largest growth occurring in urban major conurbations (6.9%) (Government Office for Science, 2021).

3.4. Religion and Infirmary

The fact that 35.0% of the town's residents listed "Undecided" for their religion suggests that they do not follow any religion at all. According to the data, 33.3% of people identify as Christians, 18.2% as Catholics, 9.6% as Methodists, 2.1% as Muslims, and the remaining people identify as Jews and Sikhs as shown in Figure 3.

Figure 3: Religion Practices in the town

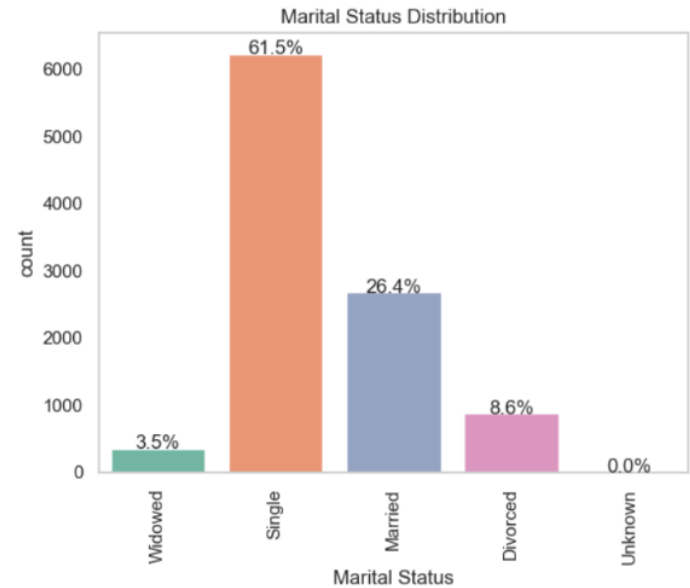


3.5. Divorce and Marriage

There have been 1,336 marriages in the town, making 2,672 persons married. With a marriage rate per thousand of 264.08. There are 871 divorced persons in the town overall, or

86.08 divorces per thousand. Data estimate shows that the number of annulments has been decreasing, with 231 cases in 2021, down from 257 in 2020 and 297 in 2019 (Office for National Statistics, 2021).

Figure 4: Marital Status in the town



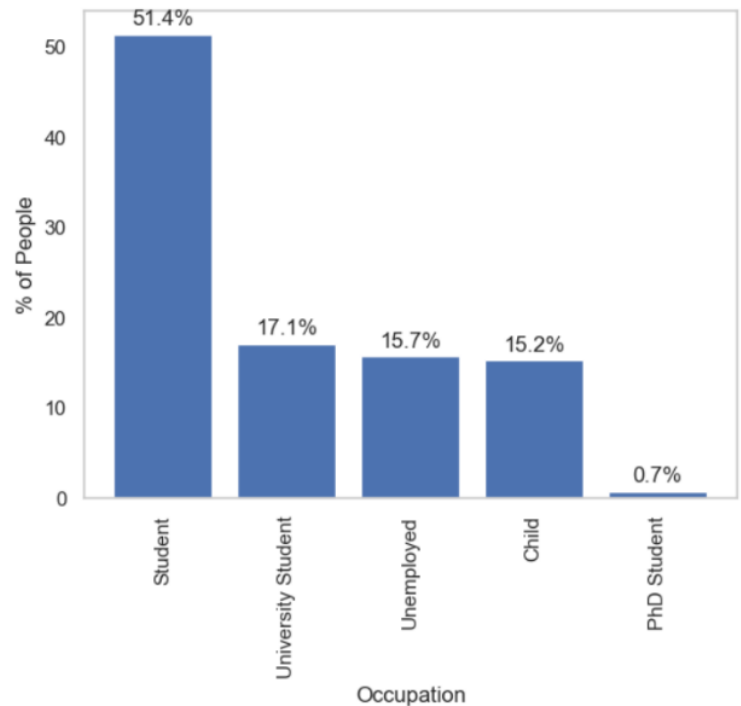
3.6. Occupancy Rates

The town estimated population of 22,048 homes was determined by accumulating each house's unique number and the number of streets in the town. The median occupancy is 3 it shows that median number of people living in the house is 3 people.

3.7. Employment

15.7% of the town's population are unemployed. Additionally, 17.1% are University students, 0.7% are PhD students, while 51.4% are student as shown in Figure 5.

Figure 5: Occupation Groups



3.8. Commuters

669 persons in all, including 27 PhD candidates and university students (the town does not have a university), are thought to commute. This indicates that commuters make up an estimated 52% of the town's population. The following methods were used to determine the number of commuters:

- a. All PhD and university students were regarded as commuters.
- b. Associate Professor Job was extracted from the data to be the commuters of the town.

4. Recommendations

The section below contains the details of the suggestion as well as the logic that lies underlying it for the government. The information from the brief about pertinent factors to take into account during the decision-making process is shown in the Table 1 below.

Table 1 : Brief criteria for making decisions

Category	Points to Consider	Decision
A Land Plot	In the event that population growth is substantial.	High-density housing
A Land Plot	If there is a need for large family dwellings and the population is considered "affluent"	Low-density housing
A Land Plot	Building a rail station could relieve traffic on the town's highways if there are a lot of commuters	Train station
A Land Plot	In the town, there is already one Catholic house of worship. Is a second church needed, and if so, of what denomination?	Religious building
A Land Plot	Not an entire hospital, rather ¹ centre for small injuries. This should be built if there are a lot of injuries or anticipated future pregnancies in the community.	Emergency medical building
Investment	Should there be proof of high unemployment, individuals ought to receive new skill training	Employment and training
Investment	Additional funds for end-of-life care will need to be set aside by the municipality if there is indication that the number of retired individuals will rise in the coming years.	Old age care
Investment	Spending on education should rise If there is evidence ¹ of an increase in the number of school-age children (new infants, or families moving to the town), then spending on education should go up.	Increase spending for schooling
Investment	Services (trash collection, road maintenance, etc.) will need greater funding if the town is growing	¹ General infrastructure

Source: project brief

4.1. What should be built on an unoccupied plot of land that the local government wishes to develop?

The process of decision making will be modelled as elimination by aspects approach whereby the number of choices is reduced based on the outcomes of the analysis gradually and at the last the final decision is reached. The relevant options include:

1. **Considerable rise in population:** Population growth though, it is at a gradual pace of about 1.0%. The low rate of growth and negative growth rate does not suggest a high population which will mark a short-term in its options hence, it will be eliminated.
2. **Demand for large family housing:** High-paying employments are a sign of being affluent. Given that the median number of occupants in a house is three, while the maximum number of people living is 22, the demand for large families is reflected in the affluent nature of the community. Still, commuters make up the majority of the town's population. As a result, large family dwelling will no longer be an option.
3. **Train station:** If the number of commuters is large: Since the majority of people living in the town are typically regular users of the roads, keep this option.
- 1 4. **Place of Prayer and worship:** Select this option because the data indicates that, of those who practice a religion (aside from the value None for religion), about one-third are Christians. However, they do not currently have a place of worship, in contrast to the 18% of Catholics who do.
5. **Emergency Healthcare Construction:** The natural increase rate and relatively low birth rate do not indicate a strong correlation with a notable increase in pregnancies in the near future, despite the 1.0% population growth rate. In fact, only a very small part of the population has any sort of disability. As a result, this option will be dismissed.

To sum up, there are two possible outcomes: either a train station or a house of worship can be constructed. In each scenario, there are respectable substitutes, such as the Catholic Church and the roadways currently travelled by commuters. Given that Christians make up 33.3% of the population and commuters make up 52% of the population, creating a rail station would benefit the majority of people more. Therefore, the final choice I will suggest is building a train station.

4.2. Which one of the following options should be invested in?

Here, the identical procedure from the previous section will be used, and it is as follows:

1. **Employment and training:** The town's unemployment rate is 15.7% of the total population, according to estimates. This choice will remain with me due to the elevated rate of unemployment.
2. **Old age care:** Approximately 9.8% of people are 65 years of age or older, and 6.8% of people are 70 years of age or older. The percentage of people 75 years of age and older continues to decline, with 3.1% of the population falling into this age group. Even if elder care is crucial, the evidence does not point to a significant proportion of the population living longer. As a result, this choice will be removed.
3. **Increase spending for schooling:** It appears that every child in the town is enrolled in school because most of the youngster in the town has the occupation student. As a result, this choice will be removed.
4. **General infrastructure investment:** Even though the town is growing at a 1.0% average pace, it cannot be said to be growing very quickly. As a result, there is no indication that the infrastructure and systems in use today are under stress. As a result, this choice will be removed.

Out of the four options I reviewed only one was kept implying that the other three were discounted and this is the one which retains people for a new set of skills for I need to find a new job. In conclusion, two solutions are going to be implemented comprising of the construction of the train station for relaxation of the traffic gridlock. Investing in transportation improvements and reskilling the people for new skills should be considered when addressing the issue of the elevated unemployment rate.

5. Bibliography

Department for Environment Food & Rural Affairs (2021) *Rural population and migration* (2021) GOV.UK. Available at: [Rural population and migration - GOV.UK \(www.gov.uk\)](https://www.gov.uk/rural-population-and-migration) (Accessed: 20 April 2024).

Government Office for Science(2021) *Trend Deck 2021: Urbanisation* (no date) GOV.UK. Available at: [Trend Deck 2021: Urbanisation - GOV.UK \(www.gov.uk\)](https://www.gov.uk/trend-deck-2021-urbanisation) (Accessed: 20 April 2024).

Office for National Statistics (2021) *Divorces in England and Wales - Office for National Statistics* (2021). Available at: [Divorces in England and Wales - Office for National Statistics \(ons.gov.uk\)](https://ons.gov.uk/divorces-in-england-and-wales) (Accessed: 20 April 2024).

Office for National Statistics (2020) *National population projections, fertility assumptions: 2020-based interim - Office for National Statistics* (no date). Available at: [National population projections, fertility assumptions: 2020-based interim - Office for National Statistics](https://ons.gov.uk/population-projections-fertility-assumptions-2020-based-interim) (Accessed: 20 April 2024).

Social Mobility Commission(2020) *Moving out to move on: understanding the link between migration, disadvantage and social mobility* (2020) GOV.UK. Available at: [Moving out to move on: understanding the link between migration, disadvantage and social mobility - GOV.UK \(www.gov.uk\)](https://www.gov.uk/moving-out-to-move-on) (Accessed: 20 April 2024).

Project_Report-5.docx

ORIGINALITY REPORT

8%

SIMILARITY INDEX

0%

INTERNET SOURCES

1%

PUBLICATIONS

8%

STUDENT PAPERS

PRIMARY SOURCES

1

Submitted to University of Hull

Student Paper

6%

2

Submitted to Unitek College, LLC

Student Paper

1%

3

cy.ons.gov.uk

Internet Source

<1%

4

Submitted to University of Wales, Bangor

Student Paper

<1%

5

Submitted to Herzing University

Student Paper

<1%

Exclude quotes On

Exclude matches Off

Exclude bibliography On