How does Land-Use affect where we retire?

Emily Winter 06/06/2023

Executive Summary

The population of England, like many other countries around the world, is ageing as people are both living longer and having fewer children. This trend looks set to continue and will likely come to influence many of the decisions nations take in the future, especially those relating to the communities where people choose to retire.

This investigation aims to discover the answers to two questions:

1. Where do people in England choose to retire?
2. What if any Land-Use features affect this decision?

This investigation uses two datasets: the first is census data from 2021 containing the population age demographics for the 309 Local Authority Regions in England. The second is Land-Use data from 2018 that describes what the land is used for in each of the 309 regions. The data was prepared and stored in memory using the Pandas library for python using its dataframe tools for manipulation and analysis.

The Regions were divided into High and Low retiree demographic areas, before a k-means classifier was used to find clusters in the various subsets of Land-Use features that matched the retiree divide. This approach was able to find a subset with a cosine similarity coefficient of 0.8489. The subset itself was one containing the Land-Use features: 'Forest, open land and water' and 'Outdoor recreation'.

This led me to inspect how these features correlate with retiree population numbers, and found that retiree population was positively correlated with 'Forest, open land and water', while negatively correlated with 'Outdoor recreation'. Indicating that retirees prefer open more wild and natural regions, so long as those areas are not then converted to become recreational spaces for the masses. This could indicate that as we get older we learn to appreciate unspoilt areas of our country more than when we are younger.

This investigation has some limitations, most significantly, the absence of historical data showing how these trends change over time. There is no accounting for population movement to show where people previously lived and no data for how areas have had their land use changed to suit this movement of people. Additionally performing this kind of demographic analysis could lead to policymaker bias if they later make decisions based on this kind of data.

However, I believe these kinds of investigations will help us better prepare and make more informed decisions as our population continues to age in the decades to come.

(Executive summary: xxx words)

Aims and Objectives

This study investigates the land use features common to areas that have high retiree populations in England. It combines two datasets: Census data detailing population demographics across England's Local Authority areas, and Land Use statistics showing how different land uses are distributed across England. The census data is from 2021 for England and Wales, and the Land-use data covers only England and is from 2018.

This study considers two questions: what areas of England have the highest population of retirees, and what land use features contribute to these areas being popular retirement destinations?

This report does not require any specialised knowledge or understanding.

Background

Much of the developed world is experiencing a demographic shift; Life expectancies are increasing and birth rates are falling, leading to an overall ageing population. By 2030, '1 in 6 people in the world will be aged 60 years or over.' (World Health Organisation, 2022), And in England alone in 2022, 19% of the total population were over 65 (Centre for Ageing Better, 2022). This ageing demographic presents unique challenges and opportunities for urban planners, policymakers, and communities alike; as understanding where older people reside is crucial for developing targeted policies, allocating resources efficiently, and providing appropriate services to meet the country's evolving needs.

This investigation aims to shed light on the importance of knowing where older people live in England and explore the specific land-use features that they gravitate towards. It uses census data alongside another dataset that describes land use data for local authority areas in England. I have previously investigated this latter dataset (Winter, 2023) but it does not contain any information on population demographics of these local authorities. This investigation combines the two datasets, to see how land-use correlates with population demographics.

One limitation of this investigation is that they are static in time, and do not show how things have changed over the years. For example they don't show population movement or how local authority land use has evolved as populations migrate.

Additionally I do not believe the Land-use statistics will have changed significantly in the years between 2018 and 2021 to make the time difference between the two datasets of concern for this investigation, in part due to the Global pandemic occurring during this time likely halting most ongoing developments although I cannot verify this.

Data sources

The Land-Use data used for this investigation is available from the Department for Levelling Up, Housing and Communities and the Ministry of Housing, Communities & Local Government of the UK government. The Census data is available from the UK Office for National Statistics. Both datasets used for this investigation are licensed for reuse under the Open Government Licence v3.0, which states that I am free to:

- "copy, publish, distribute and transmit the Information;
- adapt the Information;
- exploit the Information commercially and non-commercially."

provided that I:

- "where possible, provide a link to [the] licence."

All the work in this report is allowed under this licence. Further details and links to the data sources are given in Appendix 2.

Neither of these datasets contains any identifiable information about individual people. However performing demographic analysis on areas of the country does have the potential to lead to biassed or discriminatory policy decision making, for example if a policymaker knows an area contains one particular demographic that doesn't contribute to their interests, they might choose to focus their efforts or funding on another. Both of these datasets are publicly available and various types of demographic analysis are already public record minimising any ethical risk caused by this investigation alone.

Land-Use Data

The Land-Use data was provided to me by the Open University for use in this report but is otherwise obtainable from the UK government website. The Land Use statistics are derived from data produced by Ordnance Survey Ltd and assessed against standards set by the UK Statistics Authority; the Data Quality concern rated low as the Ordnance data was subsequently checked by the Ministry departments own team. (Ministry of Housing, Communities & Local Government, 2020)

The dataset is presented as part of a larger Excel document. The relevant page, Table P400b, Contains the total land area (hectares) by usage type for England, English Regions and English local authorities for 2018. Within this table each row contains the breakdown of Land use across multiple categories for a single Local Authority Area or Region.

Census Data

The Census data was provided to me by the Open University for use in this report but is otherwise obtainable from the Office for National Statistics website. This census data is deemed to be of high quality with 97% of households responding, further iInformation on the quality and collection methodology of UK Census data can be found online (Office for National Statistics, 2023). There are some issues with rounding errors in this dataset due to it rounding populations to the nearest 100 people, but these errors will not affect this investigation as they are insignificant compared to the total population numbers.

The dataset is again presented as part of a larger Excel document. This investigation is interested in page P02, which contains the resident population numbers separated by five-year age group for local authorities in England and Wales from 2021.

Investigation Workflow

All analysis done in this investigation is documented in the notebook EMA_project_diary (Winter, 2023), some analysis was repeated from that present in the notebook TMA 02 (Winter, 2023).

Census data import and preparation

This data was cleaned and imported into a DataFrame using the Pandas Library for Python. The entry for each local authority area was then adjusted so each one contained the population above 65 years old and the number below this age. 65 was picked as historically this was the state pension age and is the closest bracket to the current state pension age. Additionally no divide is taken into account for state pension age men and women as they are equal as of 2018 (Institute for Fiscal Studies, 2023).

The original dataset contained entries for 375 Local Authority Regions in England and Wales. The Entries for Wales were removed so this investigation can focus on England as there is no data for Wales in the Land-Use dataset leaving just the 309 Regions for England.

The prepared dataset has the entries for each of the Local Authority Regions and their Population totals, both retirees and non-retirees.

Classifying high retirement regions

In order to discover which areas in England have higher concentrations of Retirees first I calculated the percentage of the population in each Local Authority Areas that were over 65. As this allowed for differences in Region size and overall population to be mitigated. As a region with numerically the largest population of Retirees may not necessarily have the highest proportionally if it also has a very high number of people under the retirement age.

These region retiree percentages are summarised below using a Choropleth map.

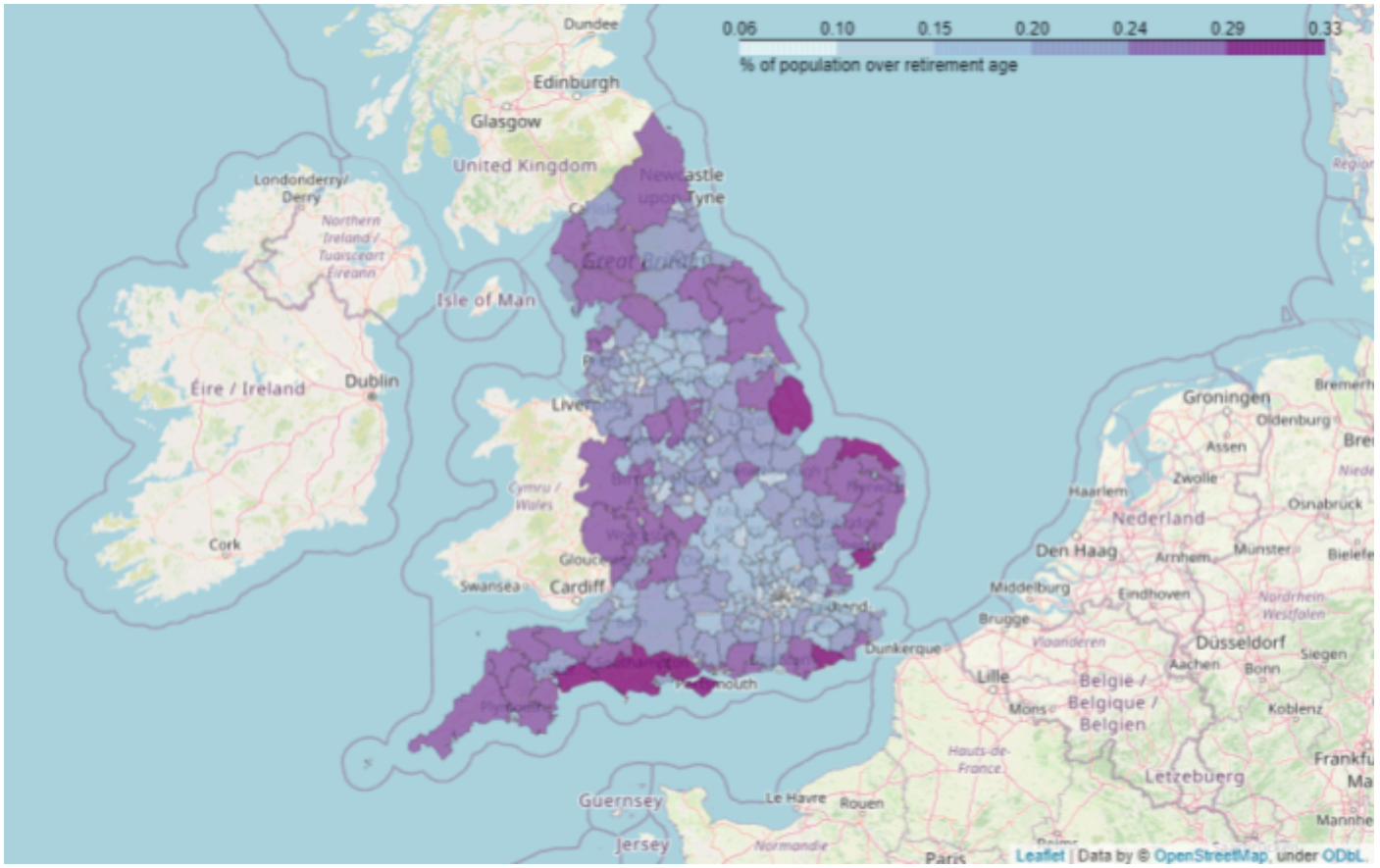

Figure 1: Map showing percentage of population over 65 years of age by Local Authority Region

With these Region Retiree percentages I then explored what metric to use in order to decide what makes a 'High Retiree' region. I plotted 4 different metrics to see how that divided up the data set as shown in Figure 2.

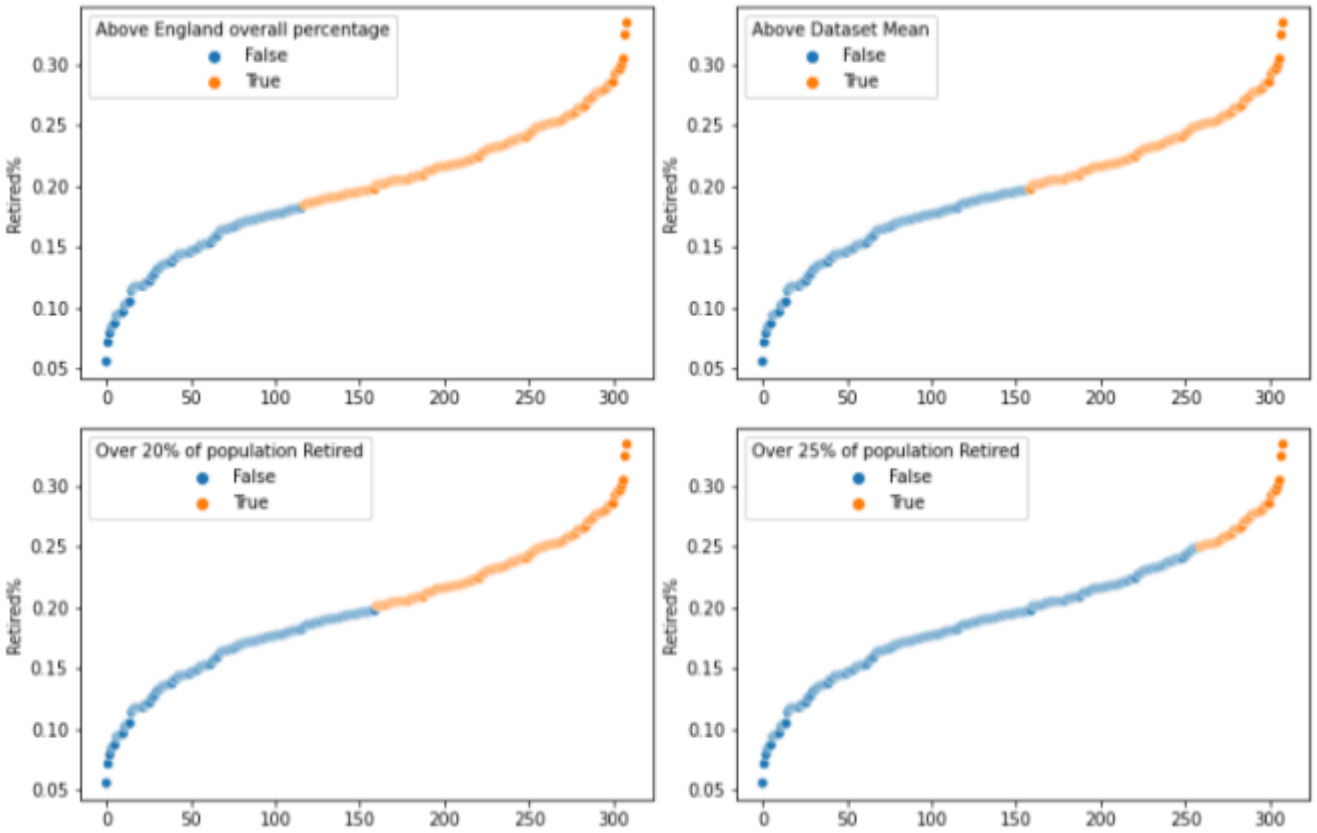

Figure 2: Comparing High Retirement Metrics.

In the end, I decided to use Metric 4, which classifies regions with over 25% of their population aged 65 years and above as 'High Retiree' regions. I had two main reasons for choosing this approach: Firstly, the United Nations designates countries as 'Super Aged' when their proportion of individuals aged 65 or above exceeds 20% (Independent Evaluation Group, 2019). Therefore, if certain areas significantly surpass this threshold, it could be a cause for concern. And secondly, this division provided a logical split for the dataset, resulting in a reasonable number of regions in both the High and Low retirees clusters.

Land-Use data import and preparation

This data was cleaned and imported into a DataFrame using the Pandas Library for Python. The details of preparing this data are similar to those described in TMA 02 (Winter, 2023) so any steps common to that investigation and this one won't be repeated. However, as the Census data used in this investigation is from 2021 and the Land-Use data is from 2018, additional preparation was needed. Specifically, several Local Authority areas had merged or changed names in the years between 2018 and 2021 so the Land-Use entries for these regions were combined to form their up-to-date counterparts replacing the originals.

The prepared dataset contains entries for the 309 Local Authority Regions in England each with 38 Land-Use categories.

What Land-Use features are prevalent in high retirement regions

The aim here was to try and discover sets of Land-Use features that created similar groupings of Local Authority Regions to the two groups created by splitting out the 'High Retiree' Regions.

To start, the Land-Use statistics were converted from Acres to a percentage of the Regions total area, then the total number of different Land-Use categories was reduced inorder to focus the scope of the investigation as well as to reduce the total number of subsets needed for future steps. These remaining Land-Use features in the dataset were Normalised, in order to further minimise the effect that differing region sizes has on the data.

Next a K-means classifier was trained on each possible subset of these Land-Use categories to split each subset into two clusters. I then used two different similarity coefficients, Cosine and Jaccard, as well as a simple percentage match test to find the clusters that most resembled the 'High Retiree' cluster created earlier. These steps were done in order to mine the data for any Land-Use groups that can be split into two clusters resembling the high and low retiree division created earlier.

The Land-Use clusters with the highest similarity were then inspected to find trends in the types of Land-Uses present.

Findings

Classifying high retirement regions

In this investigation I looked to determine what regions in England have the greatest concentrations of older people, specifically those aged 65 and over. The Local Authority Regions with over 25% of their population aged 65 and over are presented below in Figure3.

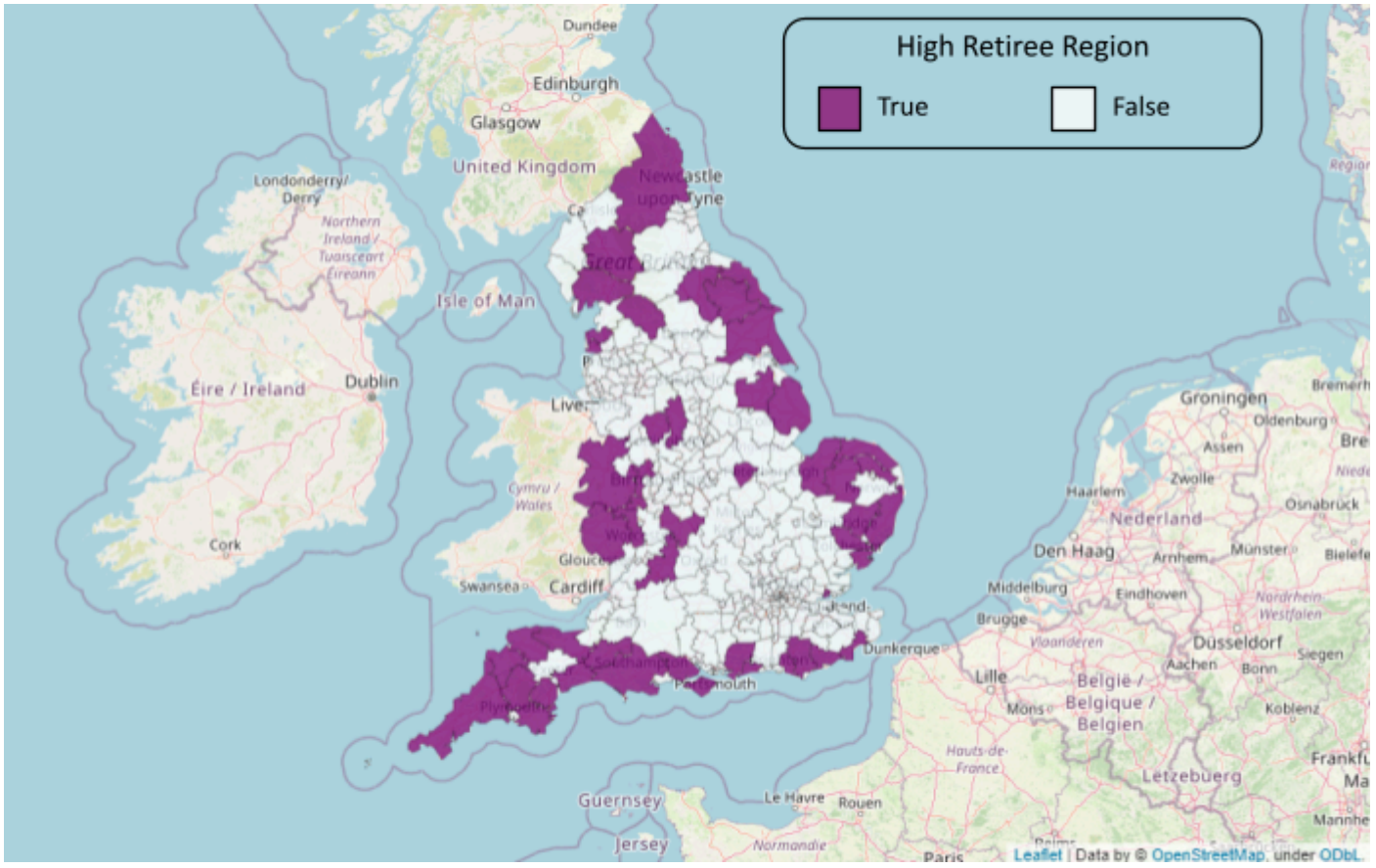

Figure 3: Local Authority Regions in England where at least 25% of the population are aged 65 and over

Of the 309 Local Authority Regions in England, 50 have these incredibly high Retiree populations. On a visual inspection, the high retiree regions are predominantly coastal and on the borders of the country away from major cities. The 10 regions with the highest proportion of retirees are listed in Table 1.

Local Authority Region	Population under 65	Population 65 and over	Total Population	Percentage of population over 65
North Norfolk	68,500	34,400	102,900	33.43%
Rother	62,900	30,200	93,100	32.44%
East Lindsey	98,900	43,300	142,200	30.45%
East Devon	105,600	45,200	150,800	29.97%
Tendring	104,400	43,900	148,300	29.60%
Dorset	267,300	112,300	379,600	29.58%
New Forest	124,000	51,600	175,600	29.38%
Isle of Wight	99,500	41,100	140,600	29.23%
Torridge	48,800	19,500	68,300	28.55%
West Devon	40,700	16,200	56,900	28.47%

Table 1: Retiree demographics for the 10 regions with highest proportion of retirees

What Land-Use features are prevalent in high retirement regions

After repeating the clustering multiple times and then calculating the different similarity coefficients on the results, I found that the cosine similarity coefficient gave the most consistent results. I discovered that the Land-Use features that consistently feature in the most similar clusters were 'Forest, open land and water' and 'Outdoor recreation'. The set consisting of just those two features, once split into two clusters using the K-means classifier, created the most similar grouping of the Local Authority Regions to the High and Low Retiree groupings with a Cosine Similarity Coefficient of 0.8489.

Once the Land-Use features most closely linked to the regional retiree population were identified, I plotted them against retiree percentage. These are shown in figures 4 & 5 below. For reference I have included colours to denote how the data was clustered, and datapoint style for High and Low retirement regions.

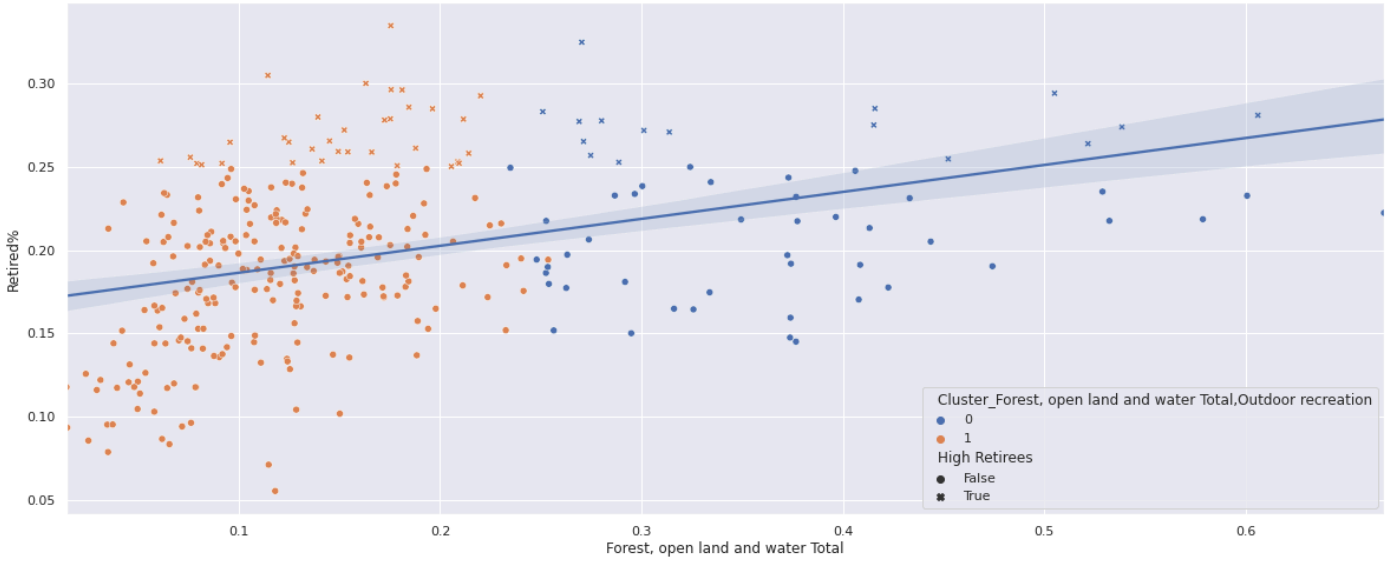

Figure 4: Retiree percentage against Forest, open land and water Land-Use

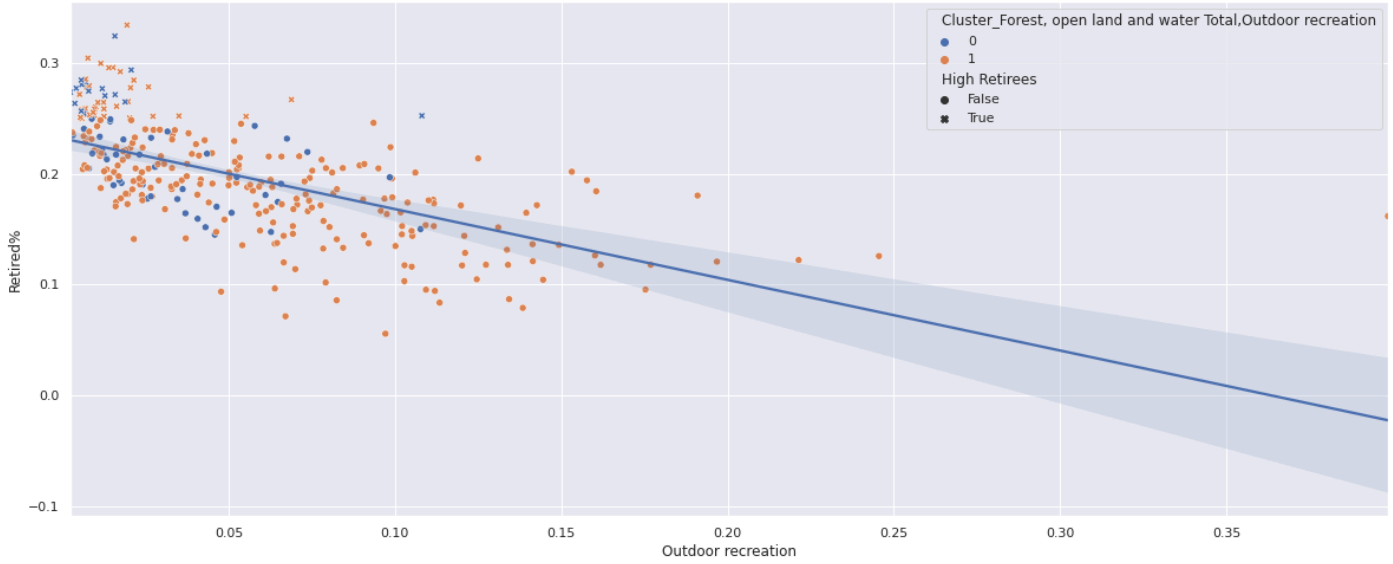

Figure 5: Retiree percentage against Outdoor recreation Land-Use

Conclusions

Investigating how Land-Use correlates to population demographics can reveal interesting relationships. In this instance it seems some of the most relevant indicators of whether people are likely to retire to an area, are if it has enough open unspoiled wild areas as indicated by the correlation between Retirees and the 'Forest, open land and water' Land-Use category. But also that said outdoor spaces are not converted into more recreational Areas, with the inverse correlation between Retirees and 'Outdoor recreation' Land-Use.

This investigation therefore, seems to demonstrate that when designating land for future development, it might be important to protect wild spaces and minimise the conversion of our natural landscape into recreational or possible tourist centric hotspots as this may come at the expense of the comfort of our retirees.

However, the usefulness of these conclusions is severely limited. Without historical data demonstrating the growth and migration patterns of England's population over time, along with the adaptation or changes in its land use to

accommodate its population; it becomes impossible to accurately determine whether the land uses attracted the retiree's to relocate there, or if they had always resided in those areas with unchanging land use features, or if the land use evolved to suit the long-standing residents who would have inhabited the region regardless.

But this investigation does start to highlight how just taking account of one feature and comparing it to another might not always be the best way to look for a pattern. As using just open land or outdoor recreation Land-Use figures to divide regions up didn't produce the most similar groupings to the retiree metric. It was only when they were taken together in a set that they grouped the regions up similarly.

Technical Reflection

I stored all data in pandas dataframes as I find these easier to work with and manipulate than a MongoDB database. Additionally the datasets themselves are small enough to fit in memory, and suit the translation from spreadsheet to dataframe due to their straight forward structure.

The datasets were of excellent quality with no missing entries and the issues of Local Authority Regions changing time the time between datasets was easily rectified. One issue I did encounter was how to create all the subsets ready for clustering. The exponential nature of subsets meant I couldn't feasibly include all of the Land-Use categories as the total number of subsets to then analyse using k-means quickly got out of hand. I had to decide what features to ignore in my analysis based on a combination of my prior experience of the dataset in TMA02 and intuition. It's likely that some of the features ignored may have been relevant but that would require further investigation using more comprehensive tools.

On reflection, I am not certain using a k-means classifier and similarity coefficients are the most suitable ways of finding the kinds of feature correlations I ended up searching for in this investigation. Other correlation discovery methods might have yielded better results, but I thought it seemed like a novel and interesting approach to mine through all subsets of Land-Use features to find if any by chance lined up nicely with the distribution of retirees in the country in this way.

(Total: xxxx words, including Executive Summary)

References

World Health Organisation (2022) Ageing and health. Available at: https://www.who.int/news-room/fact-sheets/detail/ageing-and-health (Accessed: 27 May 2023).

Centre for Ageing Better (2022) Summary | The State of Ageing 2022. Available at: https://ageing-better.org.uk/summary-state-ageing-2022 (Accessed: 27 May 2023).

Ministry of Housing, Communities & Local Government (2020) Land Use in England, 2018. Official Statistics Release. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/900910/Land_Use_in_England__2018_-_Statistical_Release.pdf (Accessed: 29 May 2023).

Office for National Statistics (2023) Quality and methodology information (QMI) for Census 2021. Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/methodologies/qualityandmethodologyinformationqmiforcensus2021 (Accessed: 30 May 2023).

Independent Evaluation Group (2019) Approach Paper: World Bank Support to Aging Countries. Available at: https://ieg.worldbankgroup.org/sites/default/files/Data/reports/ap_agingcountries.pdf (Accessed: 29 May 2023).

Institute for Fiscal Studies (2023) The planned increase in the state pension age from 67 to 68. Available at: https://ifs.org.uk/articles/planned-increase-state-pension-age-67-68 (Accessed: 30 May 2023).

Appendix 1: Notebooks

Full details of the data processing and analysis are given in the following Notebook:

EMA_project_diary.ipynb

The following notebooks containing the analysis performed in TMA02 are provided for reference:

Z8790641_project_diary.ipynb

Z8790641_q1_TMA02.ipynb

Z8790641_q2a_lab_notebook.ipynb

Appendix 2: Data catalogue

All data used in this report is licensed under the Open Government Licence v3.0, linked at: https://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/ (Accessed: 29 May 2023).

The Land Use in England data for 2018 is available from: https://www.gov.uk/government/statistical-data-sets/live-tables-on-land-use (Accessed: 29 May 2023).

The Census Population and household in England and Wales for 2021 is available from: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/populationandhouseholdestimatesenglandandwalescensus2021 (Accessed: 29 May 2023).

The digital vector boundaries for Local Authority Districts in Great Britain for December 2021 is available from: https://geoportal.statistics.gov.uk/maps/local-authority-districts-december-2021-gb-buc-1 (Accessed: 29 May 2023).