**Question 1 (45)**

See feedback comments in notebook (suffix MARKED).

1. Documentation (/5)

2. Store in MongoDB (/13)

(/1) Connect to MongoDB and create a suitably named collection.

(/4) Include the information in the hierarchical index

(/2) Correctly importing the values

(/3) Investigation of missing values.

(/3) Document structure

3. Using geographical data (/7)

(/3) Basic map

(/2) Calculating proportion of the developed land

(/2) Recognising that not all local authorities in the GEOJSON file are in the land use data

4. Converting to a relation database (/12)

(/3) Identifying suitable tables.

(/2) Appropriate primary keys.

(/2) Appropriate foreign keys.

(/5) Discussion and demonstrating that the constraints are supported by the data.

5. Database models (/8)

(/6) Two advantages and one disadvantage for each model

(/2) Clear statement of which database will be used.

**Question 2 (30)**

See feedback comments in notebook (suffix MARKED).

(a) Investigation (/20)

(/2) Basic commands to investigate

(/8) Plots and data investigation

(/2) Plot titles and axis

(/3) Folium map

(/5) Critical evaluation

-3 marks deducted if work not commented

QUESTION 2B

**REPORT: INVESTIGATING LAND USE BY ENGLISH LOCAL AUTHORITIES**

AIMS AND OBJECTIVES

The aim of this investigation was to explore land usage (in hectares) by English Local Authority Districts (LADs). Specifically, I sought to determine whether there were any geographic patterns in the distribution of different land use types across the country, and whether these patterns were associated with any particular LADs. The objectives were to identify any correlations between particular types of land use and LADs, and to evaluate the reliability and coverage of the data used in the analysis.

BACKGROUND

Land use is a critical factor in understanding the spatial distribution of economic, social, and environmental processes in the UK (specifically England for this investigation). The government has developed several initiatives to promote sustainable land use practices, including the National Planning Policy Framework (NPPF) and the Land Use Policy Group (LUPG). Understanding the patterns and drivers of land use change is essential for informing policy decisions related to urban planning, environmental conservation, and natural resource management.

SOURCES OF DATA

The data was obtained from the UK government's Open Geography Portal. Two datasets were used in the in analysis: the Land Use Statistics (LUS) dataset and the Local Authority Districts (December 2018) Boundaries UK BUC dataset. The LUS dataset contains information on the distribution of land use types across England, while the LADs dataset provides geographic boundaries for each LAD.

ANALYSIS PIPELINE

A series of data processing and analysis techniques were used. Firstly the LUS data was cleaned, transformed and stored in MongoDB. Then I used folium to visualize the geographic data and create a choropleth map of the

distribution of developed land use for Industry across England, this was done using the geojson file in combination with the MongoDB collection. I used two separate scatter plots, one to visualize the relationship between developed residential land use and non-developed natural land use, and another to visualise the relationship between land used for Industry and land use for transport.

FINDINGS

The analysis revealed several key findings. Firstly, I found that there was a clear geographic pattern in the distribution of developed land used for Industry across England, with higher values in the North of England and lower values towards the South-East. Secondly, the higher values were all located along the coast. I also found that there was a weak positive correlation between developed land used for Industry and developed land used for highways and road transport in the top 15 LADs with the highest developed land use values. There was also a weak positive correlation between developed land used for residential and non-developed natural land use values.

Text

Description automatically generated with medium confidence

Graphical user interface, text, application

Description automatically generated

*Please see L4198682\_q2a\_lab\_notebook for folium map.*

CONCLUSIONS

Overall, the investigation highlights the importance of considering the relationship between land use and geographic location in England. The findings suggest that there are variations in land use patterns across the country, which may have important implications for decisions related to urban planning, environmental conservation, and natural resource management.

I came to the conclusion that higher values for industrial land use are located primarily along the coast because, historically, ports and coastal areas have been centres of industry due to their strategic location for trade and transportation. A weak positive correlation between Industry and transport also suggests that good access to highways and road transport for the transportation of goods and services is needed in areas with a lot of Industry.

Finally, I concluded that areas with a low amount of non-developed natural land, tend to be areas with higher residential development, as seen in Birmingham and Leeds (larger cities).

REFERENCES

* Open Geography Portal. (2022). Local Authority Districts (December 2018) Boundaries UK BUC. Retrieved from <https://geoportal.statistics.gov.uk/datasets/local-authority-districts-december-2018-boundaries-uk-buc>
* Luke Bowman, LB.(2023) L4198682\_q2a\_lab\_notebook
* Open.ac.uk. (2022). The Open University. [online] Available at: https://learn2.open.ac.uk/mod/oucontent/view.php?id=2070943&section=1.2 [Accessed 4 Mar. 2023].

**593/600 WORDS**

(b) Report (/10)

(/1) Report structure

(/1) Clear and well-written

(/1) References (notebook)

(/2) Two visualisations

(/3) Critical evaluation of analysis/visualisation results

(/2) Mention of risks and uncertainties in data

**QUESTION 3A**

1. The additional dataset I have chosen is:

P02 Census 2021: Usual resident population by five-year age group. The reasons being: it contains detailed information on the population of local authorities in England and Wales broken down by age group, which can provide valuable insights into demographic trends and patterns. It will also be useful for understanding the size and distribution of different age groups within a given area. Finally, the data is highly relevant and up-to-date, and in my opinion, will be highly useful when I conduct a joint investigation with the Land use in England dataset.

1. The question I plan to investigate based on the PO2 dataset is the following. “How does the population of 18-24 year olds vary across local authorities in England and Wales? Are there any areas with particularly high or low numbers of young people?” The initial assumption would be that areas with popular Universities would have the highest number, but this investigation will allow me to confirm that, and potentially find any examples where this is not the case.
2. The question I aim to answer by combining the two datasets is the following. “Is there a correlation between the proportion of young people (aged 15-29) in a local authority area and the amount of land used for leisure and recreational buildings?”. The combination of these two datasets allows me to answer this question as they provide the necessary data. The land use dataset has data on leisure and recreational buildings, and the PO2 dataset has the necessary age group data.

**QUESTION 3B**

To address this investigative question, I will be using the P02 Census 2021 dataset, which provides population estimates by five-year age groups and local authorities in England and Wales. Specifically, I would focus on the estimates for the 18-24 age group. I will also be using the Land Use 2018 data to carry out a joint investigation using both datasets.

To store and analyse the data, I will firstly be using Open Refine for any initial data cleaning and to sort and filter it to focus on the 18-24 age group and their corresponding local authorities. I will then import the relevant data from the P02 worksheet into Jupyter Notebooks as a data frame, and then will be able to work with the data, to better understand the distribution of young people across different areas.

To visualize my results, I will likely create a choropleth map of England and Wales, with each local authority shaded according to the estimated number of 18-24 year olds living there. Additionally, I may create charts or graphs to display any patterns or trends that emerge from my analysis.

Data mining techniques may also be useful for exploring any relationships between the population of 18-24 year olds and other demographic variables in the area. I could use clustering algorithms to group local authorities with similar characteristics, or association rule mining to identify any frequent co-occurrences of different variables.

Chart

Description automatically generated**QUESTION 3C**

**Question 3 (/25)**

(a) Dataset and questions (/10)

(/2) Dataset chosen and justification

(/4) Question on new datasets

(/4) Question on both

(b) Storage and analysis plan (/5)

(/2) Storage

(/1) Folium/visualisation technique

(/1) How they will be used

(/1) Mention data mining

(c) Workplan and project diary (/10)

Workplan (/6)

(/2) Milestones

(/3) Plan relating to tasks

(/1) Realistic time scales

Project diary (/4)

(/1) Initial code cell

(/1) Discussion of combining

(/2) Further work in notebook