

Sydney Liveability Analysis

F12D-RE08 - 5

Group Members:

Mengjia Zhu SID:500627341

Jing Hu SID:490223941

Liou Liu SID:490255986

Introduction

Today, with the rapid development of global economy, people's quality of life is also improved, so whether a region is livable has become one of the important indicators when people choose to live in a community. Therefore, our group conducted a series of investigations and studies on livability scores in Greater Sydney and City of Sydney. The purpose of calculating the livability score of Greater Sydney area and Sydney areas is to have a relatively objective judgment standard when looking for a suitable living area.

In this project, we will integrate multiple datasets to calculate the livability scores of different areas in Greater Sydney, And based on this, we will choose reasonable suggestions to the stakeholder, young photography enthusiasts, and related research staff.

Dataset description

In the Sydney Liveability Analysis, two CSV files provided general information for each area. Additionally, three shape files provided school catchments information, crime rate information and boundary data, which will help in calculating the livability score. A brief description of each dataset will be provided below.

Neighbourhoods.csv

This dataset includes census data for Sydney containing information on land area, population by age and total population, accommodation/dwelling, number of businesses, and financial information such as rent and income for each area of Sydney. During the processing of the dataset, the NaN values and Null values are replaced with 0 and duplicate rows are removed to facilitate subsequent habitability calculations. In order to facilitate subsequent calculations, "young people" is added to the table as a new column, which is the sum of the population aged 0-19 in the region.

BusinessStats.csv

This dataset contains business information for various regions, which includes the number of different business types (e.g., accommodation and catering services, etc.) and information on public facilities

(profits, transportation, post, etc.). NaN values in the dataset will be replaced with 0 and duplicate rows will be removed.

School_catchments.zip

This dataset is actually used concat() function in pandas to connect “catchments_primary.shp”, “catchments_secondary.shp” and “catchments_future.shp” three datasets. And deleting the data that has the duplicate ‘use_id’ and keeping the first one. This dataset contains information about public schools in New South Wales, Australia. This dataset provides information on catchment areas, registration times and student age groups for different types of schools. Additionally, this shape data provides the size of each catchment area, providing polygons for the boundaries. SRID is GD94(4283)

Break_and_enter.zip

This dataset is shape data of theft 'hotspots' in NSW as determined by BOCSAR. This dataset provides crime rate data for different regions of New South Wales, Australia. SRID is GD94(4283)

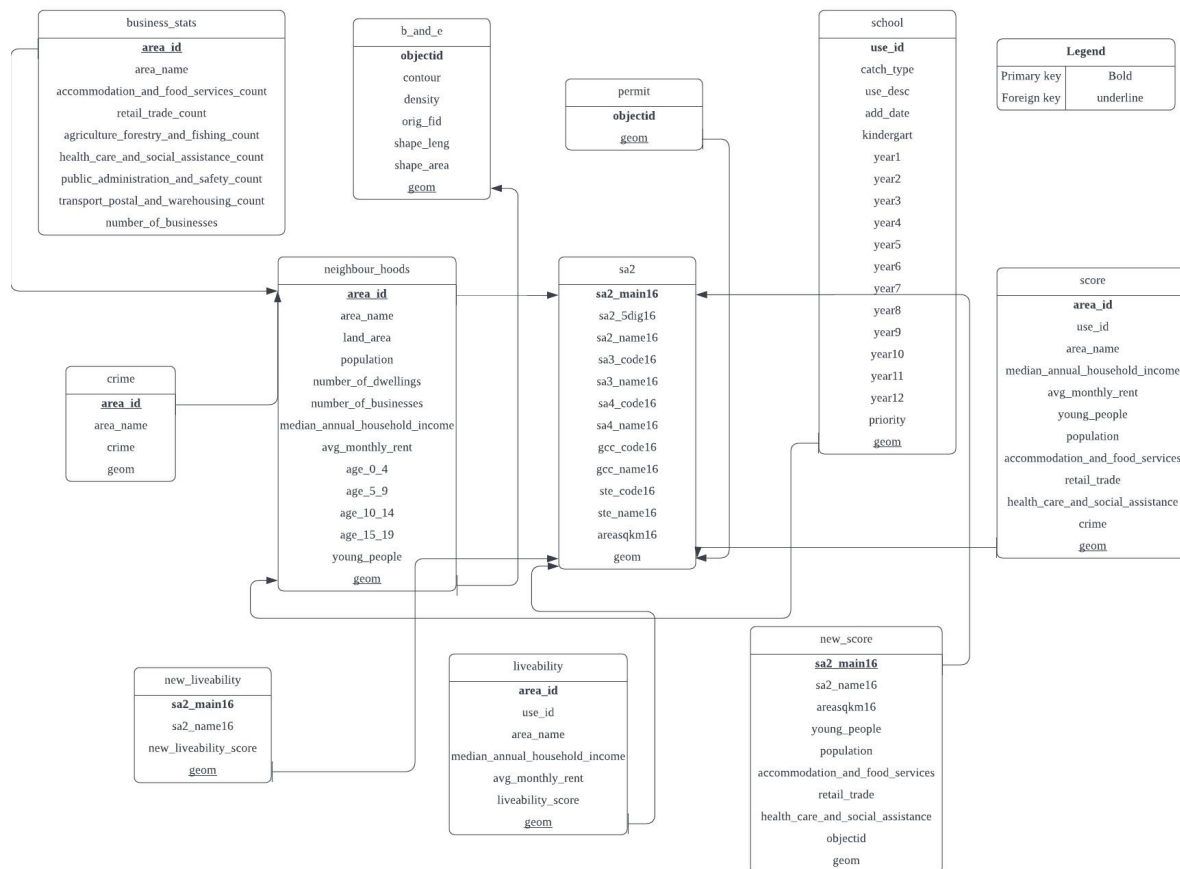
SA2_2016_AUST.zip

This dataset contains shape data about suburban areas in Australia. This dataset provides information on area codes and names and more information about the region. Additionally, this shape data provides the size of each area and polygons for the boundaries. SRID is GD94(4283)

Additional dataset - Filming permit areas.shp

This dataset provides areas in Sydney that require a filming permit, including organizations to apply for, websites and contact details.. In addition, the shape data provides the size of the various regions that need to be applied, and provides polygons for the boundaries.

Database description



The dataset contains 11 tables, namely 'business'_stats', 'b_and_e', 'permit', 'school', 'crime', 'neighbour_hoods', 'sa2', 'score', 'new_Liveability', liveability 'and' new '_score'. They contain information on the population, business, schools, crime rate and livability of Greater Sydney and the City of Sydney.

The primary key and foreign key settings of each table are shown in the above table. The primary key is marked in **bold** and the foreign key is marked with underline. In addition, we set the column 'geom' of 'SA2' and 'neighbor' respectively as an index to ensure the uniqueness of data, speed up the retrieval speed of data and the connection between tables, and pave the way for more efficient analysis in the future.

For the connection between tables, we use the 'St_ The contains' function to ensure whether the region to be queried is within a certain range, which is more inclusive and prevents omission.

Greater Sydney Score Analysis

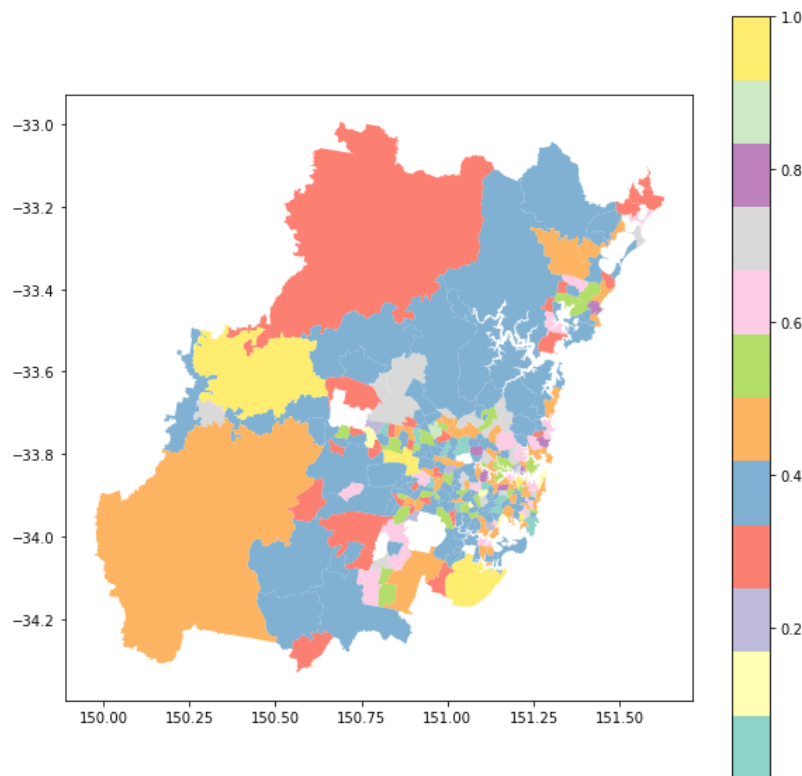
The calculation of the livability score for each area is done through panda. The formula followed in the calculation is below:

$$\text{Livability score} = S(Z_{\text{school}} + Z_{\text{accomm}} + Z_{\text{retail}} - Z_{\text{crime}} + Z_{\text{health}})$$

The sigmoid function (S) and z-score (Z) are used in this formula. Among the 5 variables, "school", "retail", "health", and "accommodation" have a positive effect on livability, and "crime" has a negative effect on habitability.

In the calculation of crime, since we consider that only high crime density will have a significant impact on the livability of an area, we only select the regions with high density during the calculation. After that, connect sa2 and neighbor_hoods using area_id, and then connect b_and_e and neighbor_hoods using geom to obtain the sum of hotspot areas and total area, then divide to get X(crime).

Create a new table for calculating school, accomm, retail and health related data and create a new column for each section of the calculation. After it, calculate the Z-score for each of the 5 sections. Finally, use the formula of calculating the livability score mentioned above to calculate the livability score of each area and store it in a new table.

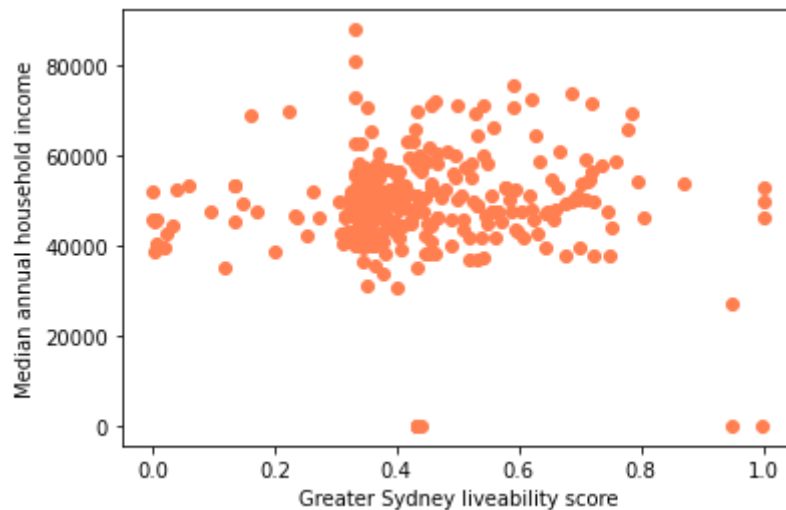


(For our map plot, the x-axis is longitude, and the y-axis is latitude)

As can be seen from the above figure, the livability of most areas in greater Sydney is about 0.3 to 0.5. In addition, in the northeast and west of the map, there are areas with high livability, and the livability scores are 0.8 and 1.0 respectively. These places are very suitable for people to live in. Besides, according to the comparison of the Median Annual Income Diagram in Appendix(1), the correlation between the regional livability score and the median annual household income is very weak.

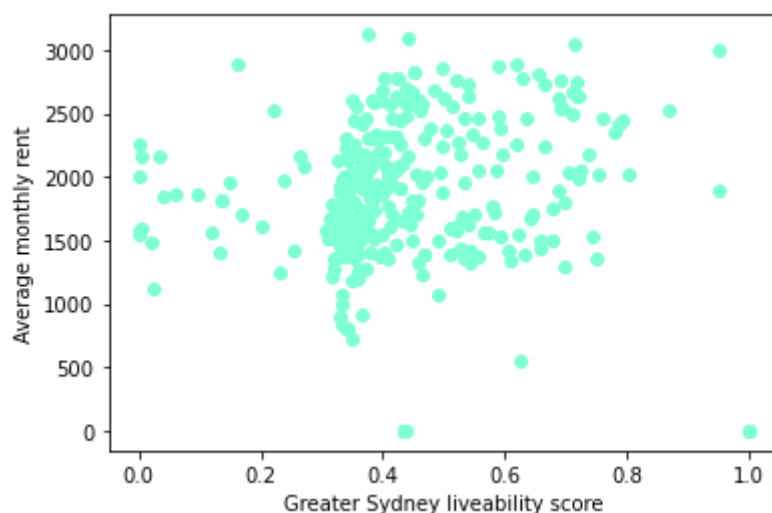
Correlation Analysis

There is a very weak correlation between median annual household income and the livability score, with a Pearson correlation coefficient of 0.006. The figure below shows the visualization results.



The reason for the extremely weak correlation may be that median annual household income is a good indicator of the income level of the population in the area, but does not correctly reflect other information about the area (such as information about facilities in the area). Due to the large geographical range of Greater Sydney, there are different people living there, including high-income and low-income people. The reasons for this phenomenon may be that people with lower incomes live in the same community with imperfect infrastructure (for example, there are no supermarkets or shopping centers around, so it is not convenient to shop, housing quality is not uniform, there are no schools in or near the community with poor teaching quality, no developed transport network, etc.). Finally, there is a very weak correlation between median annual household income and greater Sydney Livability score.

There was also a very weak correlation between the average monthly rent and the livability score, with a Pearson correlation coefficient of 0.073. The figure below shows the visualization results.



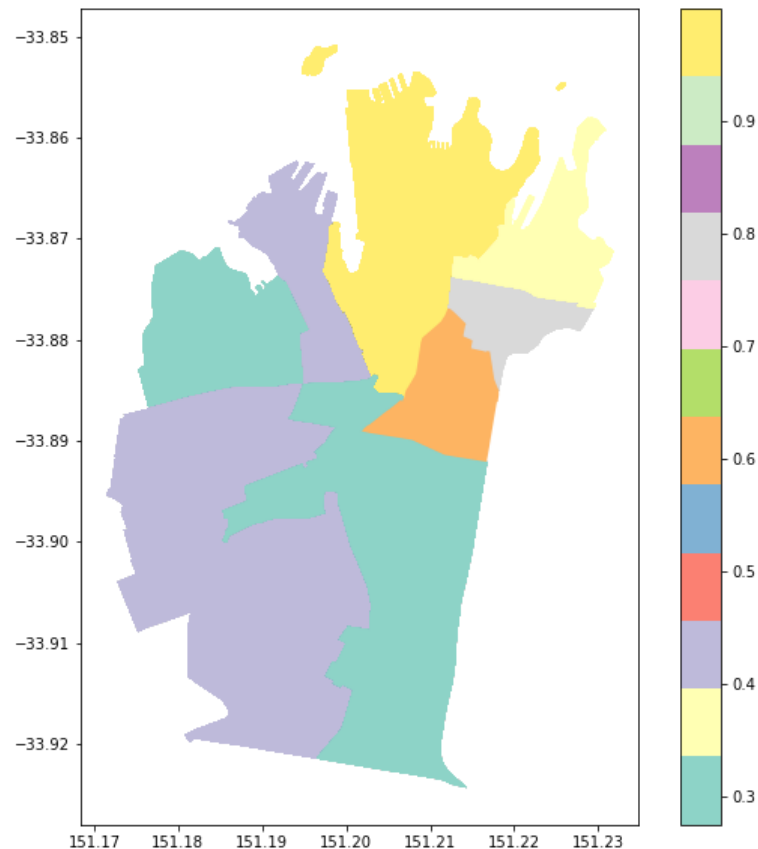
The reason for the weak correlation may be that the average monthly rent does not well represent the livability of the whole region. Because the area of greater Sydney is very wide, there is a gap between the rich and the poor. The rent of a high-end community is higher than that of an ordinary community, and its livability is generally higher than that of an ordinary community, which will have bias on the overall livability analysis.

City of Sydney Analysis

We calculated the liveability of different areas of Sydney for young photography enthusiasts using the following formula:

$$\text{Livability score} = S(Z_{\text{permit}} + Z_{\text{accomm}} + Z_{\text{retail}} + Z_{\text{health}})$$

The sigmoid function (S) and z-score (Z) are used in this formula. Because the target group is young people without children, catchment areas are not considered. In the formula, all four variables are positively correlated with the habitability score. The calculation method is the same as before. The visualization results are as follows:

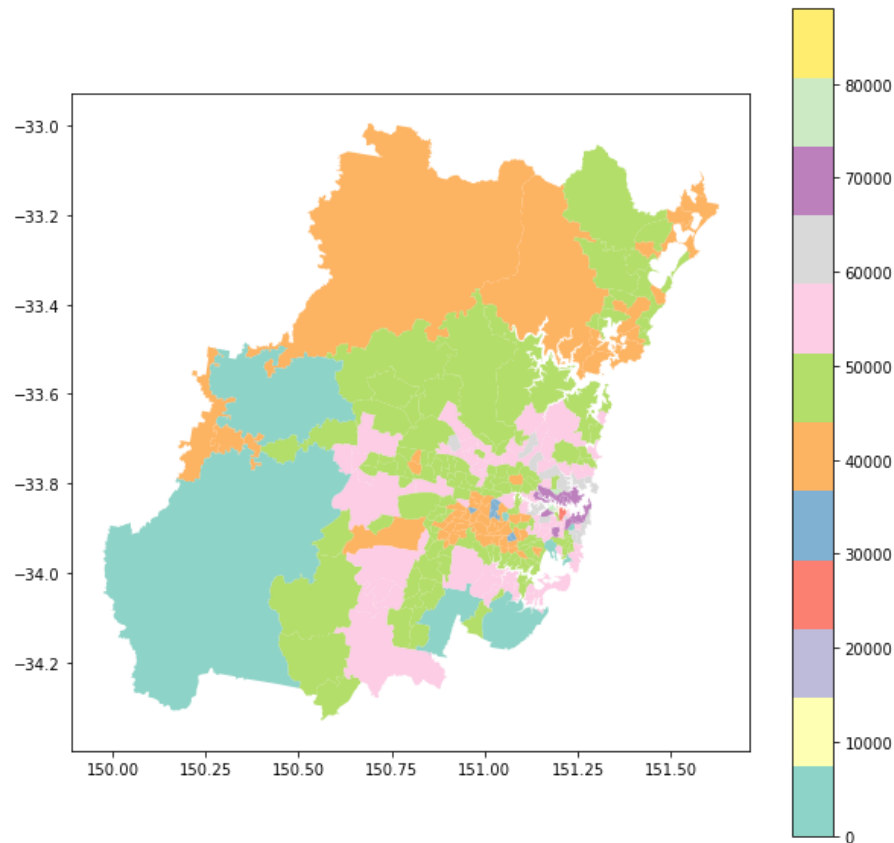


(For our map plot, the x-axis is longitude, and the y-axis is latitude)

As can be seen from the above figure, for young photographers and related research staff, the northern or central region contains more allowable shooting sites than other places, and the conditions of accommodation, retail and health are also excellent. It not only ensures a good basic living environment, but also has more space for them to carry out their interests, so the northern or central area of greater Sydney is more suitable for them to live.

Appendix

1. Median Annual Income Diagram



Reference

Dataset source

Neighbourhoods.csv

Provided by the coordinator of DATA2001, University of Sydney

BusinessStats.csv

Provided by the coordinator of DATA2001, University of Sydney

School_catchments.zip

Provided by the coordinator of DATA2001, University of Sydney

Available at:

<https://data.cese.nsw.gov.au/data/dataset/school-intake-zones-catchment-areas-for-nsw-government-schools>

Break_and_enter.zip

Provided by the coordinator of DATA2001, University of Sydney

Available at:

https://www.bocsar.nsw.gov.au/Pages/bocsar_datasets/Datasets.aspx

SA2_2016_AUST.zip

Provided by the coordinator of DATA2001, University of Sydney

Available at:

<https://www.abs.gov.au/statistics/standards/australian-statistical-geography-standard-asgs-edition-3/jul2021-jun2026/main-structure-and-greater-capital-city-statistical-areas>

Filming Permit areas.shp

Available at:

<https://data.cityofsydney.nsw.gov.au/datasets/cityofsydney::filming-permit-areas-1/about>

Others

Schema diagram

Available at:

https://lucid.app/lucidchart/75cf928c-8ccc-4425-961e-2ac77537eccc/edit?invitationId=inv_b49597d9-aa68-468f-8505-54b0072c2c0b&page=0_0#

Visual diagram

From the submitted jupyter book file