Sydney liveability score analysis

Group F08F-RE03 - 9

Introduction：

In recent years, more and more people choose to travel, live and study in Sydney, which leads them to need a guide to give them information about Sydney. We looked at the liveability of various parts of Sydney to help people who are visiting or settling down. Let them have different considerations when choosing a place to live. In this report, we will be tasked to gather and integrate several datasets in order to investigate the most 'liveable' suburb for a potential stakeholder to buy in Greater Sydney through calculation of a 'liveability' score. This report is aimed at residents, tourists, overseas students, landlords and real estate agents who need to know about housing resources. Similarly, government departments can use the report as a reference to confirm future urban planning and prevent potential hazards such as robbery and theft.

Dataset Description：

BusinessStats.csv：This datasets has 2309 rows and 9 columns. This data set quantifies how livable an area is. Examples are number of businesses and accommodation and food services. By quantifying these indicators, this data set can be used to help us analyze how livable the area is.

Neighbourhoods.csv: This dataset has 322 rows and 13 columns. The data set packets contain some basic conditions in different regions, such as occupying land and population. It also contains information about residents, such as monthly income, rent and the number of minors.

school_catchments.zip：This data set we used three files, which are the "catchments_primary.shp""catchments_secondary.shp"and "catchments_future.shp". The catchments_primary. shp file contains 1666 rows and 19 columns of data, The catchments_secondary.shp file contains 453 rows and 19 columns of data and The catchments_primary. shp file contains 44 rows and 18 columns of data. The files store geographic information about Sydney and the age distribution of the students.

break_and_enter.zip：The file that we use is the "breakenter-ing_jantodec2021.shp", contains 2594 rows and 7 columns. From this file we can obtain information about

geographical location and crime frequency.

SA2_2016_AUST.zip：The file that we use is the " SA2_2016_AUST.shp", contains 312 rows and 13 columns. This file provides map information of the whole of Australia. In the data clearing stage, we will focus on selecting the part of Sydney to help us conduct analysis.

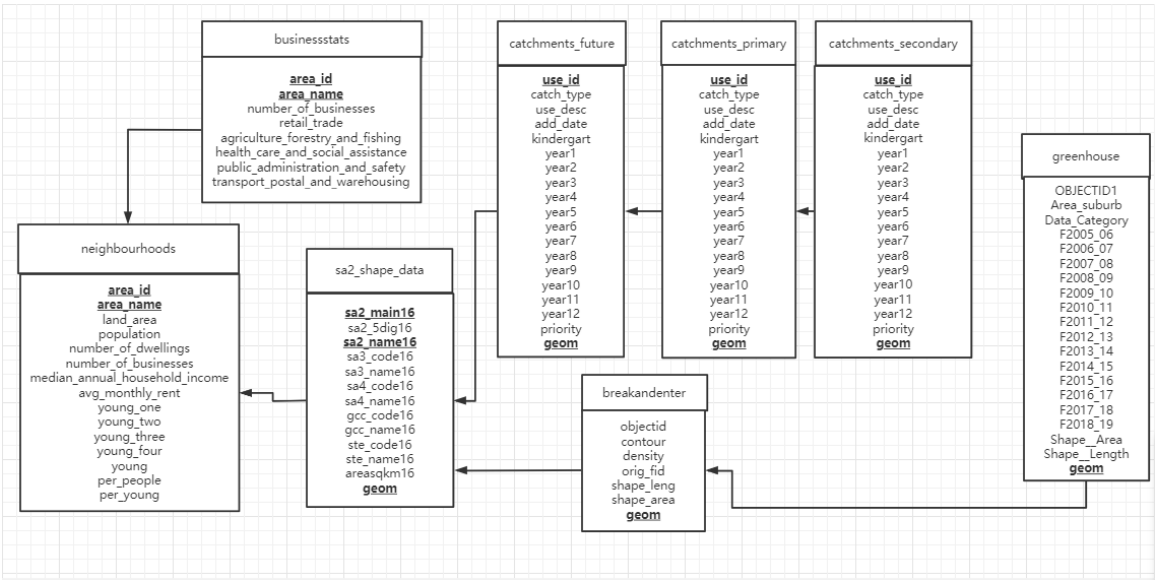
Greenhouse_gas_emissions_profile_by_suburb.geojson：This dataset has 174 rows and 20 columns. We can get information about greenhouse gas emissions in the Sydney metropolitan area at different times.("Greenhouse Gas Emissions Profile By Suburb")

Data Cleaning：

For the file “Neighbourhoods.csv”, we conducted the most detailed data cleanup to ensure that we can easily extract the most useful information. We remove the 'Unnamed: 0' column, which is likely to cause system misidentification, because it is not useful to us. This was followed by changes to the inconvenient age section, such as "0-4" to "young_one" and so on. The next step is to convert some of the values to floating-point types so that we can compare and evaluate them with other data sets. Finally, divide the 'population' and 'young' columns by 1000 to make it easier to perform the Sigmoid function in Task2.

For the document "SA2_2016_AUST. Zip", we only keep the part of 'Greater Sydney' because this report is not a study of Australia's livability situation, but only focuses on Sydney.

Database Description：



The database is about the liveability survey of Sydney. This database is based on the data set described earlier in the report. The files used are: “BusinessStats.csv”, “Neighbourhoods.csv”,“school_catchments.zip”,“break_and_enter.zip”and “SA2_2016_AUST.zip”. We associate different datasets with "geom", "area_id", and "area_name". In the neighborhoods where we'll dataset land_area set up for the index.

Sydney Livability Analyze：

Formula: Score = S(z(school) + z(accom) + z(retail) – z(crime) + z(health) –z(gas))

At the same time, we used two functions: S---sigmoid function, z --- z–score

There are six variables in this formula, and they are school, accommodation, retail, crime, health and gas emission. Among them, school, accommodation, retail, health and health all contribute to improving the livability index, while crime and gas emissions reduce the livability index. And the reason why we don't do it directly is because the units and even orders of magnitude of each variable are not in the same standard. So we used Z-score to ensure the fairness of these variables. This function allows all elements involved in the habitability index calculation to be normalize. In summary, the reason we use these two functions is that we need to distribute elements between 0 and 1 so that the computation works.

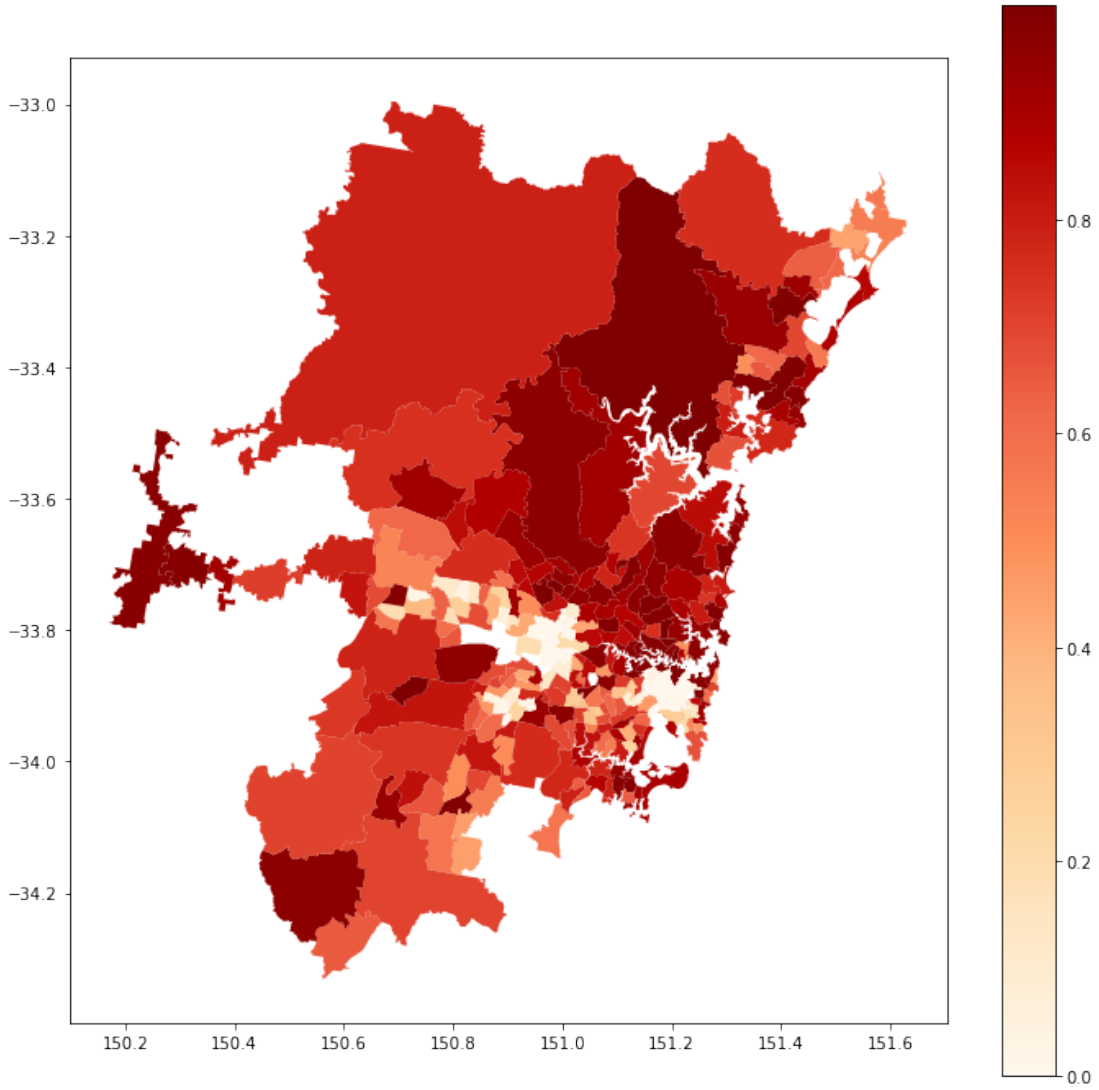


Figure1 Sydney Livability Score Visualization

The figure above shows our final result. The area with red color is more suitable for people to live in, while the area with white color is less suitable for people to live in. There were some counterintuitive color patterns, such as a white in central Sydney, which was considered uninhabitable, while some of Sydney's surrounding areas appeared quite dark red. Our guess is that high population density leads to higher crime rates and higher gas emissions, leading to a lower livability index.

Correlation Analyze：

Household Income

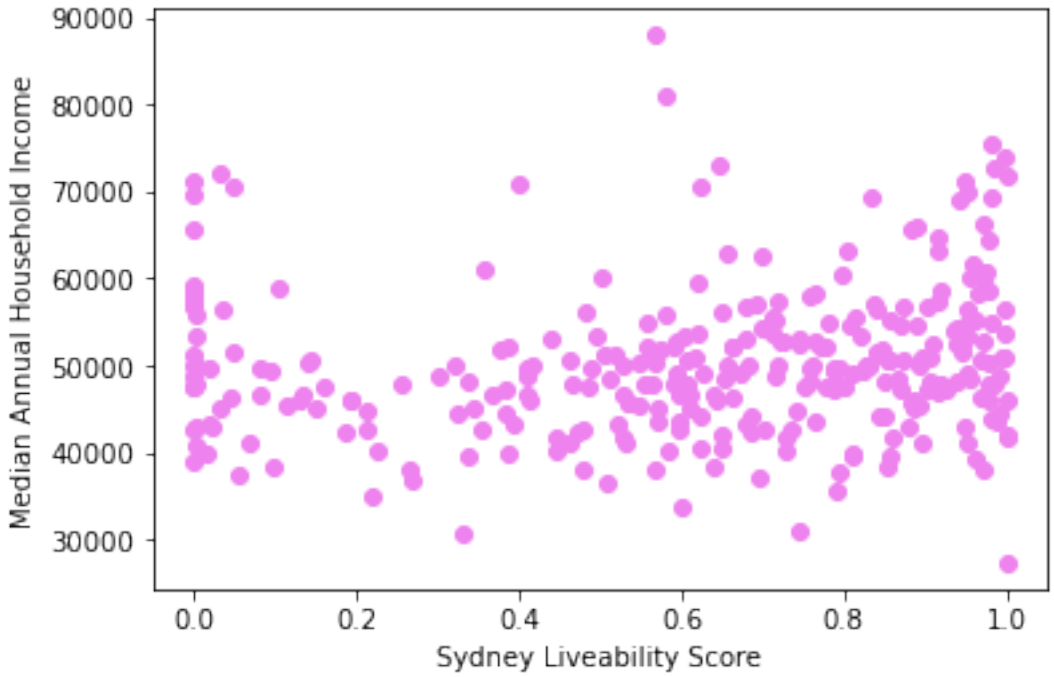


Figure2 correlation between Household Income and Sydney Liveability Score

The correlation between Sydney Liveability Score and Median Annual Household Income is 0.126. From this number we can see that there is no correlation or only a very weak correlation.

Avg Monthly rent

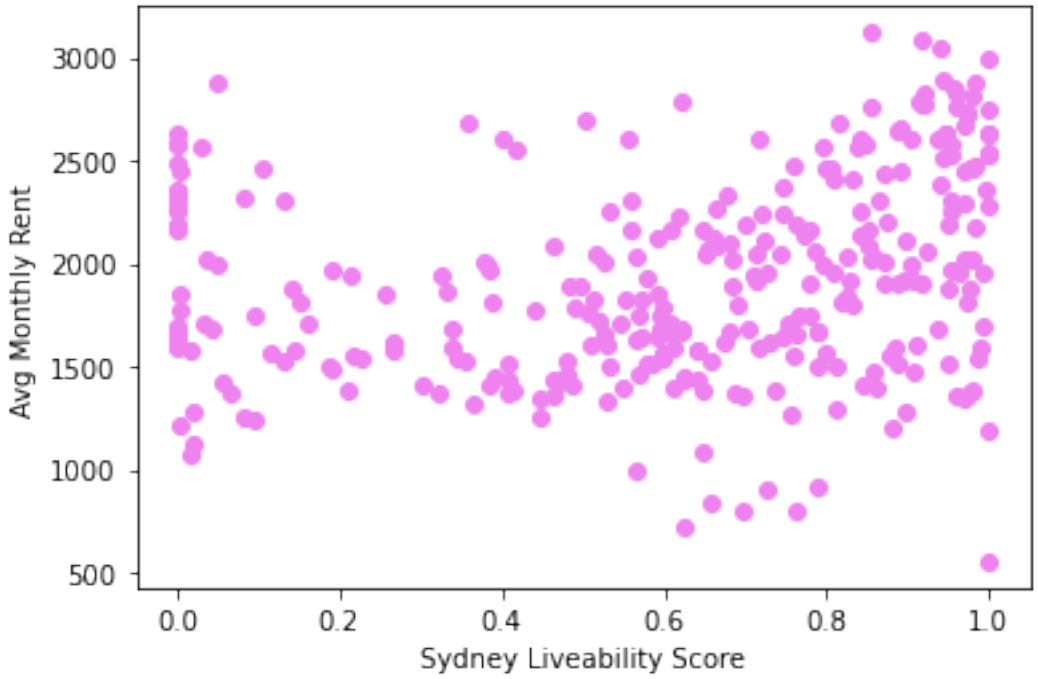


Figure3 correlation between Avg Monthly Rent and Sydney Liveability Score.

The correlation between Avenge Monthly Rent and Sydney Liveability Score is0.230. Similarly, Avenge Monthly Rent and Sydney Liveability Score exhibit no correlation or only a very weak correlation. This suggests that livability scores have little to do with residents' income or monthly rent.

Avg Monthly rent and household income

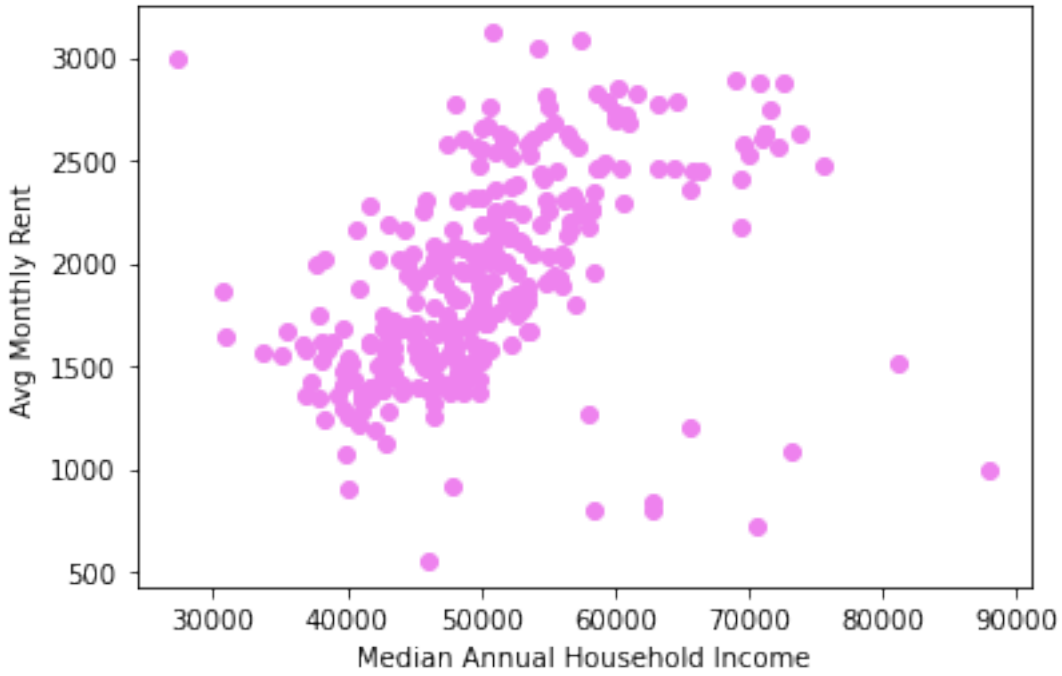


Figure4 correlation between Avg Monthly Rent and Household Income

The correlation between Avg Monthly Rent and Household Income is 0.452. There was a low to moderate correlation between the two. That means there is a correlation between income and rent, but not strong.

The higher a person's income, the more they pay in rent. And the higher the income, the less data. We think the reason for this is that high-income people choose to buy instead of renting, that would explain why the data is more concentrated in the lower and medial income segments.

Conclusion:

Through this project, we combined a total of 6 datasets into a database, and calculated the livable scores of Sydney districts by sigmoid function and Z-Score. Formula gives satisfactory results, but our analysis still has many limitations. For example, in the correlation study, we did not find data with stronger correlation, and the relevant charts were not intuitive enough. These problems will be further improved in the future.

Reference：

1,"Greenhouse Gas Emissions Profile By Suburb". Data.Cityofsydney.Nsw.Gov.Au, 2022, https://data.cityofsydney.nsw.gov.au/datasets/cityofsydney::greenhouse-gas-emissions-profile-by-suburb-1/explore.