Bushfire Risk Analysis (Report)

Dataset Description

Six datasets were used to complete our bushfire risk analysis –

1. *Statistical Areas*

This file was obtained as a .csv file. It provides information on the different regions in and around NSW, and has three columns – area_id, area_name, and parent_area_id. The parent_area_id column is a derivative of the area_id column, usually consisting of the first few digits of the area_id. It determines the larger area/region to which a smaller region belongs. Each smaller region can be uniquely identified by its area_id, and for that reason, it was chosen as the primary key.

The dataset did not contain any null or missing values, but before uploading it to the database, we converted the area_id and parent_area_id columns to object types.

2. *Neighbourhoods*

This file was obtained as a .csv file. It provides information on different neighbourhoods within the NSW region, including the name of the region, the area of the land, the population, the number of dwellings/businesses, the median income, and the average monthly rent. Each area is uniquely identified by its area_id (different from the one used in Statistical Areas), so this column was chosen as the primary key for this dataset.

Some of the columns contained missing values, and the 'median income' and 'average monthly rent' columns had the most values missing. Both were required for our correlation analysis, and the dataset itself was quite small (322 rows). Replacing the missing values with the average/median of the whole column might not have been entirely accurate, since that might have skewed our analysis, especially with such a small dataset. Therefore, we dropped all rows with missing values.

The 'population' column was stored as a string (object) type since some of the values contained commas. We removed all the commas and converted it to an integer type. The 'number of dwellings' column was also converted from a float type to an integer type. The 'number of businesses' column was dropped since that data will be obtained from the Businesses Statistics dataset.

3. *Business Statistics*

This file was obtained as a .csv file. It provides information on the different facilities available in a neighbourhood, such as the number of businesses, healthcare facilities, accommodation services, retail/trading establishments, and agriculture/forestry related establishments. Each row is identified by its area_id (with a similar format to the Neighbourhoods dataset's area_id), and that was used as the primary key for this dataset.

There were no missing or null values in this dataset and all the datatypes were appropriate, so we did not have to perform any pre-processing before uploading it to the database.

4. *RFSNSW Bushfire Risk Data*

This dataset contains spatial data on the vegetation and risk categories provided by the NSW Rural Fire Service and is formatted as an ESRI shapefile (.shp). It describes bushfire prone land in the form of Point data types and gives the area and length of these points.

The dataset did not contain any null or missing values, and all the datatypes were appropriate. Before uploading it to the database, we renamed all the columns to ensure they fit the schema.

5. *SA2 data from ABS*

This dataset contains spatial data on areas that function together as a community, socially and economically (Australian Statistical Geography Standard (ASGS)). The spatial data was stored in the form of Polygons and MultiPolygons and formatted as an ESRI shapefile (.shp).

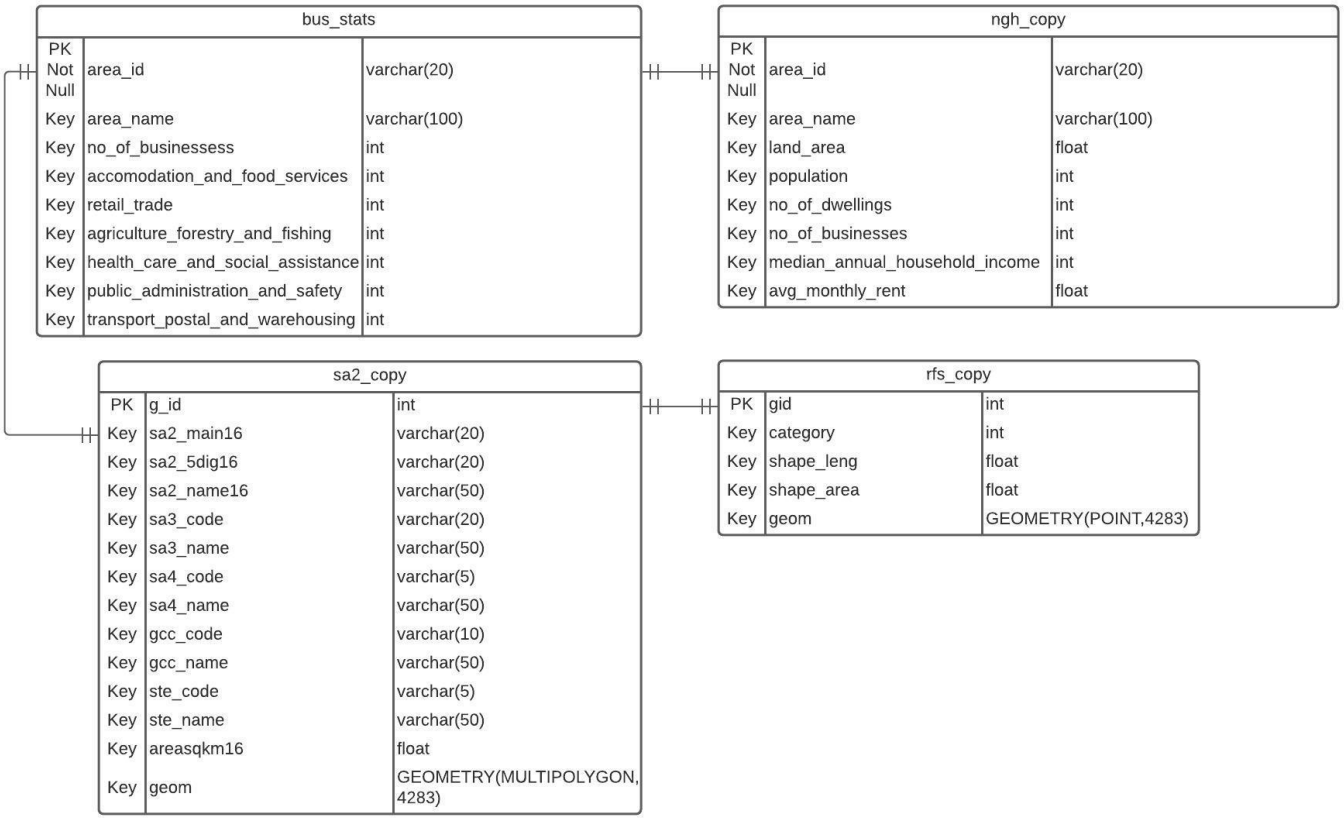
The dataset also contains information on Statistical Areas Level 3 and 4, but we are only interested in Level 2 data.

For this dataset, the spatial data (the 'geometry' column) was the most important. However, 18 values were missing from this column, and we decided to drop them so we could focus on the data we needed. Before uploading it to the database, we renamed the columns so it would fit the schema and converted the sa2_main16 column to an integer type.

Database Description

We integrated our data into the public schema on the PostgreSQL server. There were five datasets we added - Statistical Areas, Business Statistics, Neighbourhoods, NSW Rural Fire Services spatial data and SA2 data from ABS.

The Neighbourhoods (ngh_copy) and Business Statistics (bus_stats) were joined using area_id, and we referred to this dataset as merged1. The SA2 (sa2_copy) and RFSNSW (rfs_copy) datasets were joined spatially, and we referred to this dataset as sa2_rfs. The merged1 and sa2_rfs were joined using the area_id and sa2_main16 columns.



Fire Risk Score

To calculate the fire risk score, we first had to integrate all our datasets. The SA2 and RFSNSW datasets were spatially joined, and the Neighbourhoods and Business Statistics datasets were joined on the column 'area_id'. The two merged datasets were then joined on the sa2_main16/area_id columns. This gave us a dataset which described the various parameters of different neighbourhoods within NSW, including the exact positions of bushfire prone lands.

The 'category' column classifies an area based on potential bushfire risk and vegetation type. There are three categories - 1, 2, and 3. Category 1 signifies the highest risk, and is usually represented as red on a map. Category 3 stands for moderate risk and is represented as dark orange on a map. Category 2 signifies low risk and is usually represented with light orange on a map (Guide for Bushfire Prone Land Mapping, 2015). Based on the category column, we created a 'risk factor' column, assuming that Category 1 land has three times the potential risk as Category 2 land, and Category 3 land has twice the potential risk as Category 2 land.

Using this formula, we calculated the fire risk score.

$$fire\_risk\_score = S\left(z\left(\frac{population}{land\_area}\right) + z\left(\frac{number\_of\_dwellings}{land\_area}\right) + z\left(\frac{number\_of\_businesses}{land\_area}\right) + z\left(\frac{risk\_factor \times areasqkm16}{land\_area}\right) - z\left(\frac{health\_care\_and\_social\_assistance + public\_administration\_and\_safety}{land\_area}\right)\right)$$

The fire risk score is very widely distributed. The overwhelming majority of the score lies in the range -0.5 to 0.5.

Our calculated fire risk score is relative. A small change in the score can reflect a huge increase/decrease in risk. From the histogram, we can see that there is a single area which has a very high fire risk score (>15). This is a possible outlier.

	mean	median	standard_deviation
	2.042798e-16	0.176197	2.247592

The mean is very small compared to the median or the standard deviation. This shows that we have a very widely distributed score, which is also shown by the histogram.

Figure 1. Distribution of Fire Risk Score

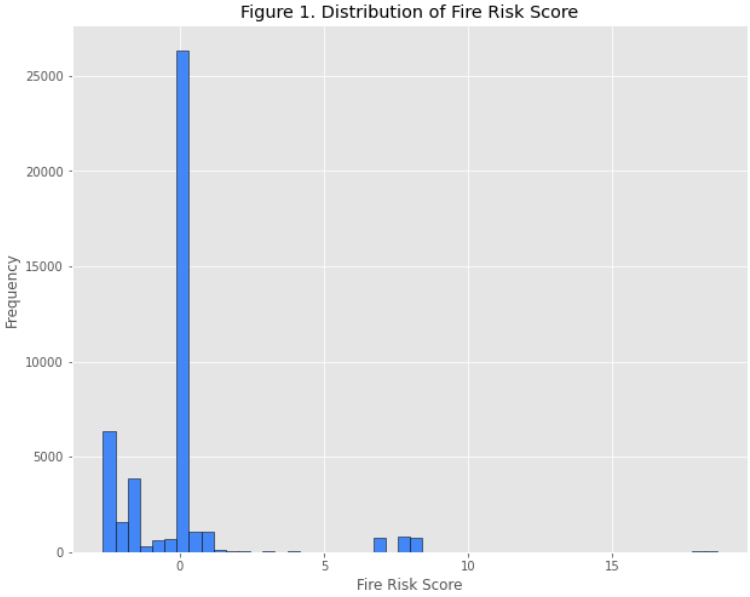
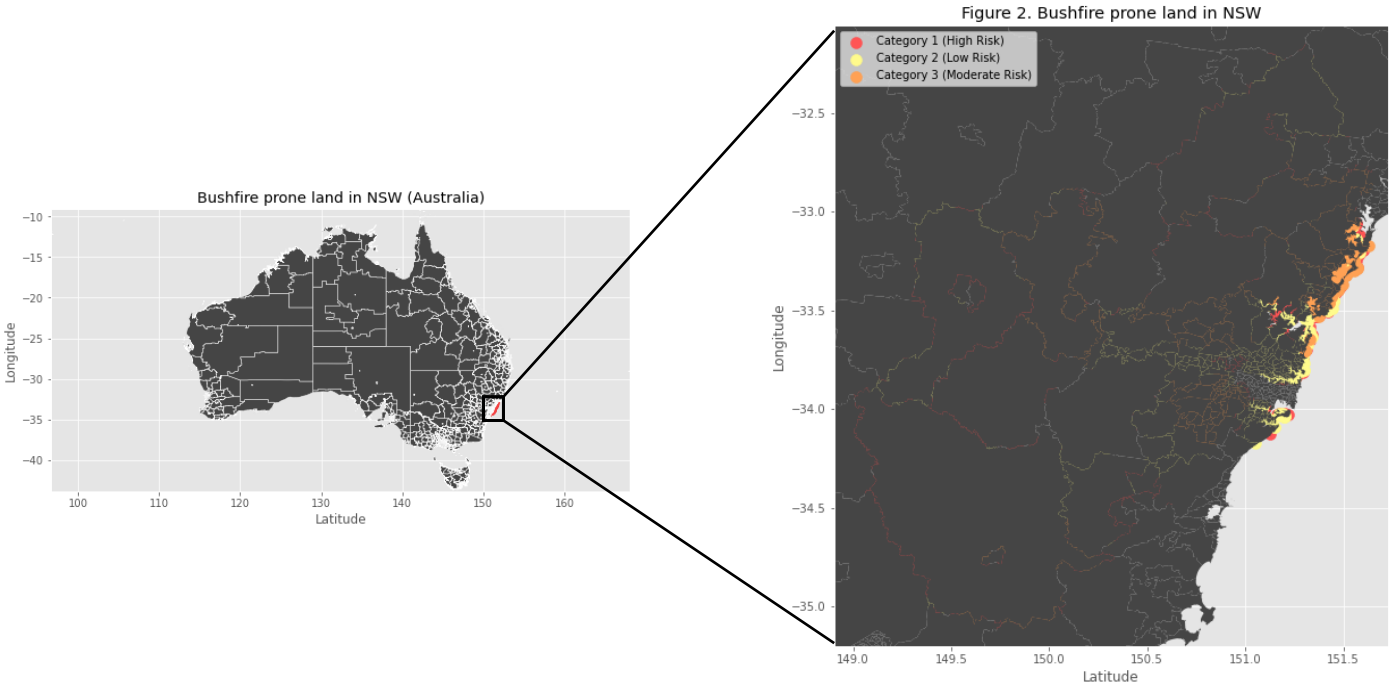


Figure 2. Bushfire prone land in NSW



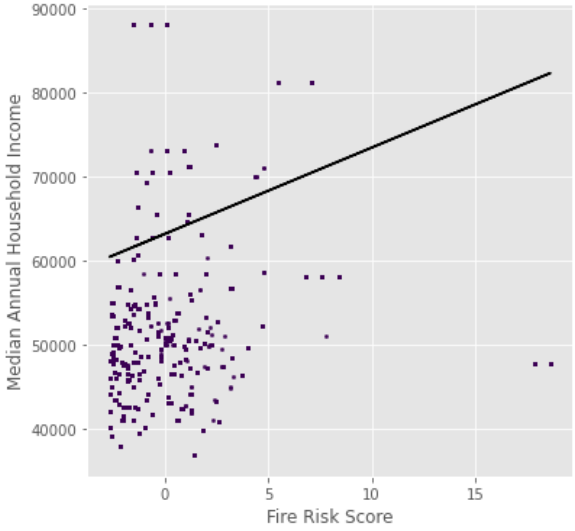
Correlation Analysis

	median_annual_household_income	avg_monthly_rent
fire risk score	0.184709	-0.167664

There were no strong correlations between fire risk score and either median annual household income, or average monthly rent, but there was a weak positive correlation between fire risk score and median annual household income, and a weak negative correlation between fire risk score and average monthly rent.

We cannot say that there is an association between fire risk score and the affluence of a neighbourhood.

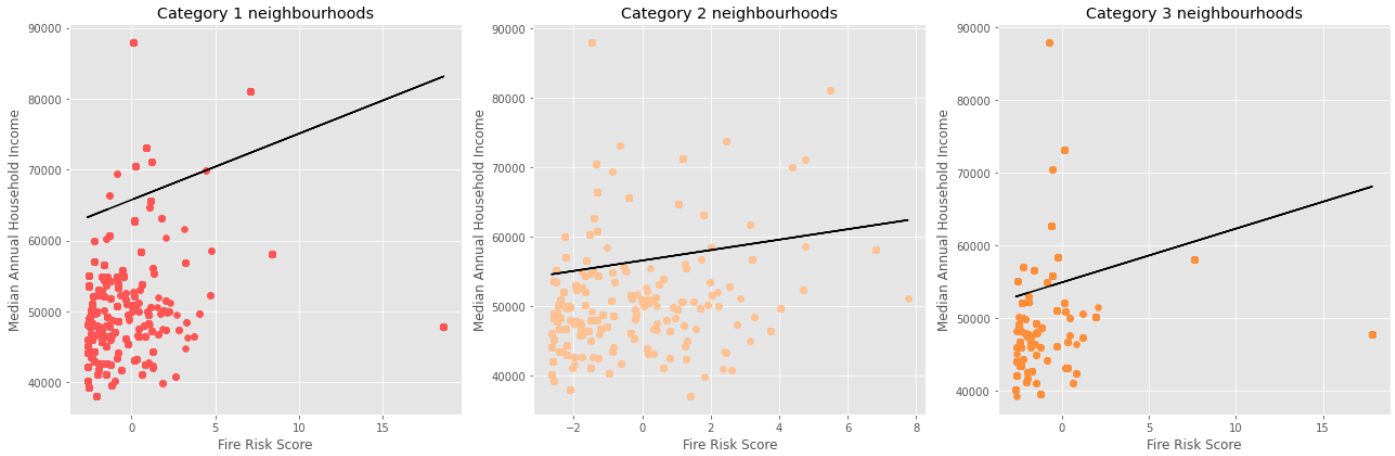
Figure 3. Correlation between Median Annual Household Income and Fire Risk Score



The graph shows the weak positive correlation between fire risk score and median annual household income.

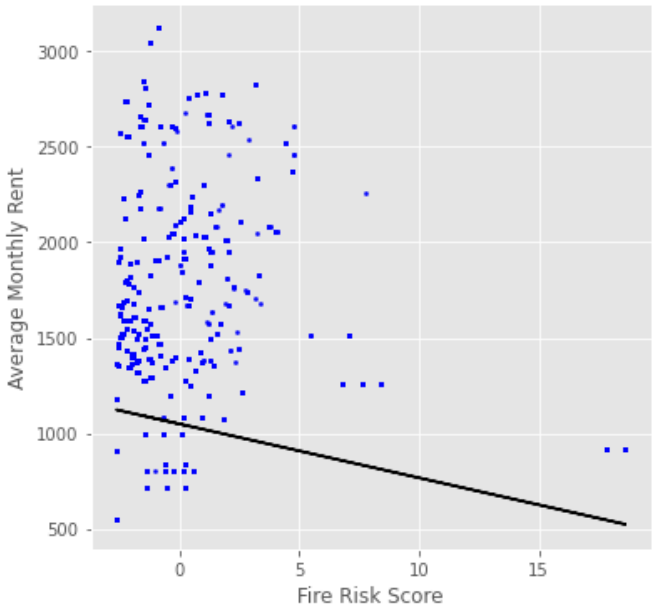
The majority of points are clustered near the lower left corner of the graph (low income, low fire risk score). There are also some points that appear to have the same median annual household income, but different fire risk scores.

We can study the correlation more by controlling the category.



Controlling for the category of the risk in the neighbourhood does not reveal any associations. The highest correlation is in Category 3 neighbourhoods, with a value of 0.29.

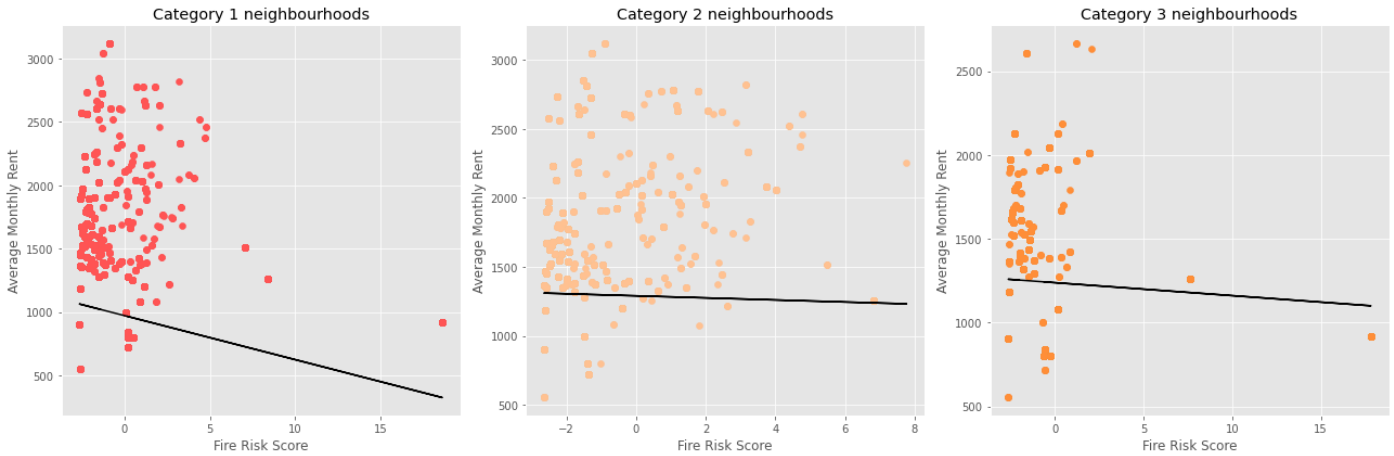
Figure 4. Correlation between Average Monthly Rent and Fire Risk Score



This graph shows the weak negative correlation between fire risk score and the average monthly rent.

The data points are more spread out in this graph, and there is a wider distribution of the average monthly rent. As before, there are many points that have the same average monthly rent, but different fire risk scores.

We can test the correlation further by controlling for category.



There seems to be no association between average monthly rent and fire risk score even when controlling for category. The strongest relationship here is a weak negative correlation (-0.19) between fire risk score and average monthly rent for Category 1 neighbourhoods.

Overall, there seems to be no association between fire risk score and affluence.

References

- Abs.gov.au. 2021. Australian Statistical Geography Standard (ASGS). [online] Available at: https://www.abs.gov.au/websitedbs/D3310114.nsf/home/Australian+Statistical+Geography+Standard+(ASGS) [Accessed 2 June 2021].
- NSW Rural Fire Service. 2015. Guide for Bushfire Prone Land Mapping. [online] p.11. Available at: https://www.rfs.nsw.gov.au/__data/assets/pdf_file/0011/4412/Guideline-for-Councils-to-Bushfire-Prone-Area-Land-Mapping.pdf [Accessed 4 June 2021].