I. Dataset Description

Our report leverages multiple online public datasets provided by different departments and agencies of the Australian government as detailed below in the following list:

- **Statistical Areas Level 2 - 2021 - Shapefile[1]:** Sourced from Australia Bureau of Statistics. It contains the boundaries of all SA2 regions in Australia represented by shapefile geometry.
- **Businesses by industry division by Statistical Area Level 2 by turnover size ranges[2]**: Sourced from Australia Bureau of Statistics (ABS). It includes the count of businesses categorised by each SA2 region, turnover size and industry sector.
- **Timetables Complete GTFS[3]**: Sourced from Transport for NSW. It contains a list of public transport stops in NSW indexed by stop id, and the location of each stop via latitude and longitude coordinates.
- **School intake zones (catchment areas) for NSW government schools[4]**: Sourced from NSW Department of Education. It contains the geographical regions that students must live in to attend a certain school. There are three groups of shapefiles for the three categories of primary, secondary and planned schools.
- **AEC - Federal Election - Polling Place (Point) 2019[5]**: Sourced from the Australian Electoral Commission. It contains a list of every voting centre created for the 2019 federal election, indexed by id and provides location of each station via latitude and longitude coordinates.
- **Population[6]**: Sourced from ABS. It includes population estimates of each SA2 region categorised by age groups and sex.
- **Income[7]**: Sourced from ABS. Includes the amount of income earners, median age of earners and the median income per SA2 from 2019-2020
- **Emissions[8]**: Sourced from SEED NSW. Includes the total amount of emissions by each LGA and is further broken down into sources of emissions such as agriculture or car fumes
- **Crime Hotspots[9]**: Sourced from NSW Bureau of Crime Statistics and Research. It contains the crime locations categorised by 11 offence types in 2022.

Pre-processing

The data preprocessing and cleaning procedures were executed within a Jupyter Notebook utilising a Python 3.11 runtime. The SA2 boundaries and school intake zones shapefiles were imported as GeoDataFrame objects using the read_file() function from the GeoPandas library. Conversely, the remaining datasets were imported as DataFrame objects employing the read_csv() function from the Pandas library.

To ensure consistency and compatibility with PostGIS, several steps were undertaken for the SA2 boundaries and school catchments datasets. The geometry column of these datasets was meticulously verified to contain entries represented in the MultiPolygon type of the Shapely library. Furthermore, these entries were converted to the Well-Known Text (WKT) format in PostGIS, utilising the Spatial Reference Identifier (SRID) 4326.

For the stops, polling places, and school intake zones datasets, a new geometry column was generated from the longitude and latitude columns using the GeoDataFrame.points_from_xy() function. Subsequently, the generated geometry column was converted to the WKT format, ensuring consistency with SRID 4326.

Several filtering and selection operations were performed on various datasets. For the SA2 boundaries dataset, only regions within the Greater Sydney Area were retained, with the SA2 code, SA2 name, and geometry columns preserved. Similarly, the businesses dataset was filtered to include only regions within the Greater Sydney Area, while selecting the industry code, industry name, SA2 code, SA2 name, and total businesses columns.

Unused columns were removed from the public transport stops dataset, retaining only the stop id, stop name, and geometry columns. In the case of the polling places dataset, rows with missing values in the longitude and latitude columns were dropped, while preserving the polling place id, polling place name, and geometry column.

To consolidate the primary school and secondary school data, the respective school intake zone datasets were concatenated into a single dataframe. The selected columns for this dataset included id, description, catchment type, and geometry.

For the population dataset, a new column called "young_people" was created by aggregating the 0-4, 5-9, 10-14, 15-19 containing the sum of the population aged from 0 to 19 . The selected columns comprised the SA2 code, SA2 name, young people, and total people.

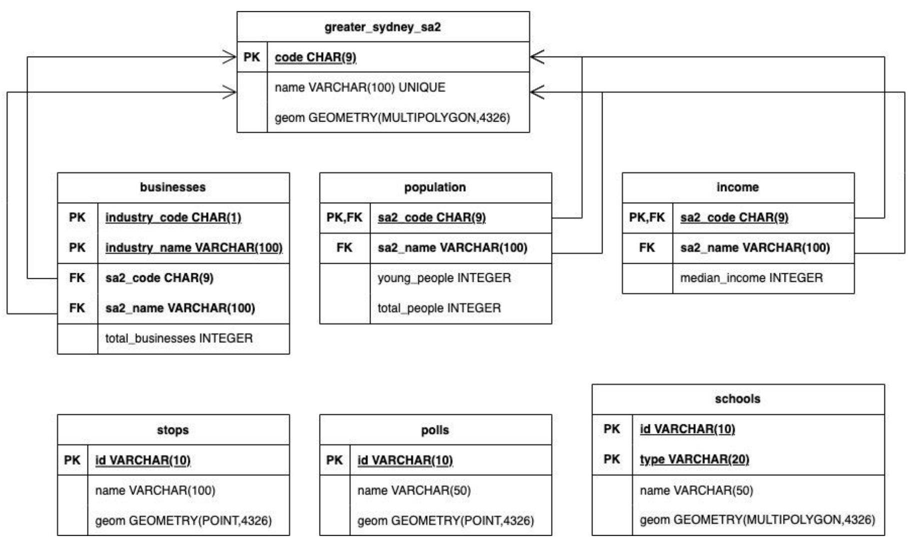
Similarly, the income dataset was filtered to encompass regions within the Greater Sydney Area, excluding rows with missing median income values. The selected columns for this dataset included the SA2 code, SA2 name, and median income.

Through these rigorous data processing steps, the datasets were refined and standardised, enabling subsequent analyses and calculations within our project.

II. Database Description

Schema Design

We employed the SQLAlchemy library to establish a connection to and conduct queries on our local PostgreSQL database from within Python. To enable spatial queries, we installed the PostGIS extension on our SQL database. To organise our datasets efficiently, we introduced a new schema named "data2901" for their storage. For each dataset, we executed the CREATE TABLE commands to establish the corresponding table structure as illustrated in the figure below. A visual representation of our database schema structure can be found below.



To seamlessly integrate the data loaded into our data frames with our database, we relied on the Pandas library's DataFrame.to_sql() function, which facilitated the insertion of dataframe data into their respective tables after we defined the schemas.

Index Optimisation

To optimise spatial join operations and facilitate the computation of scores, we created GIST indexes on the geom column of the stops, polls, and schools tables. These indexes played a crucial role in spatially joining the tables and creating views to enable score calculations by significantly increasing the execution speed of the spatial queries.

III. Results Analysis

Score Formula

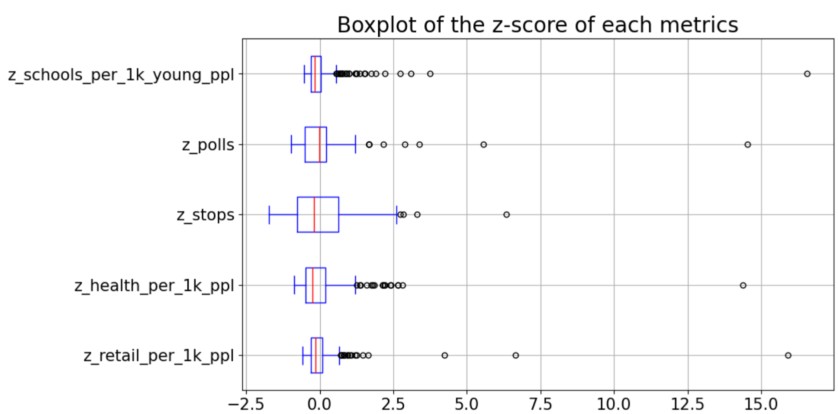
We used the following five features to compute the score for each SA2 region:

- $retail$: the number of retail businesses per 1000 people.
- $health$: the number of health services per 1000 people.
- $stops$: the total number of public transport stops.
- $polls$: the total number of polling places.
- $schools$: the number of school intake zones per 1000 young people (0~19 years old).

Since some of the features are calculated per 1000 people and because some SA2 regions are mostly unpopulated (due to factors including major infrastructure, military bases and national parks), we only included the 360 regions with more than 100 people. In the boxplot of the z-score of each feature, it is observed that there are numerous extremely large outliers, and the distributions, except for stops, are right-skewed. However, after inspecting the outliers, it has been determined that they are valid outliers and should not be removed. This is because these outliers represent the SA2 regions within the CBD of Sydney, that have significantly higher resource availability compared to the average region and are essential to include as part of Sydney.

To mitigate the impact of outliers on the score calculation, two approaches were considered: (1) applying a logarithmic transformation and (2) using robust z-scores for normalisation. After conducting several experiments, the decision was made to utilise the second approach of using robust z-scores. This was because computing z-scores using median and median absolute deviation (MAD) is more robust to outliers than using mean and standard deviation, as the distribution of scores better resembled a normal distribution compared to the z-score method, as shown in the histogram of the scores distributions below.

Each feature was normalised using the robust z-score to ensure equal impact on the final score. The z-scores are summed and passed through a sigmoid function to obtain a score ranging from 0 to 1. Therefore, the score for each SA2 region is calculated using the following formula:



$$Score = sigmoid(rz_{retail} + rz_{health} + rz_{stops} + rz_{polls} + rz_{schools}),$$

where for the i-th region in the feature x, its robust z score $rz = \frac{x_i - median(x)}{MAD(x)}$, and

the median absolute deviation $MAD = median\left(\left|x_i - median(x)\right|\right)$

Extended Score Formula

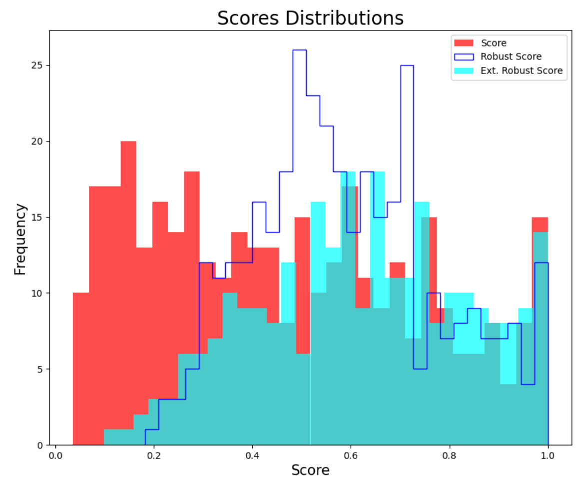
We combined the following two additional features for each SA2 region to extend our score formula:

- $emissions$: the total CO_2 emission in tons per square kilometre.
- $crime$: the number of crimes occurred per 1000 people.

After removing the regions with emissions data, we are left with 271 regions. The extended score for each SA2 region is calculated using the following formula:

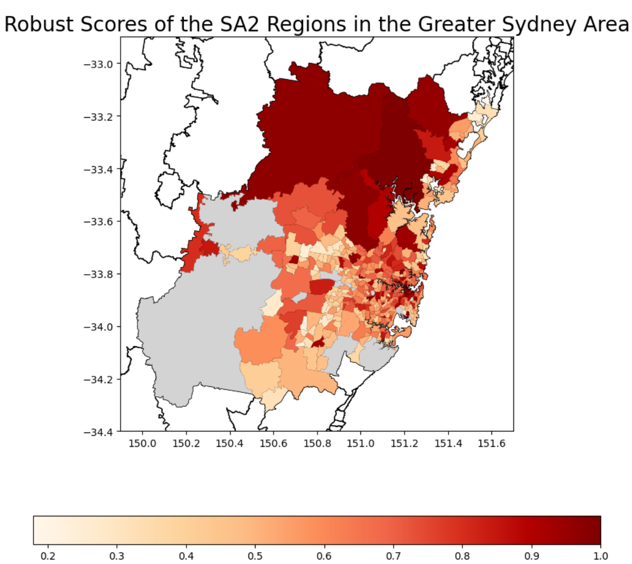
$$Score_{ext} = sigmoid(rz_{retail} + rz_{health} + rz_{stops} + rz_{polls} + rz_{schools} + rz_{emissions} + rz_{crime})$$

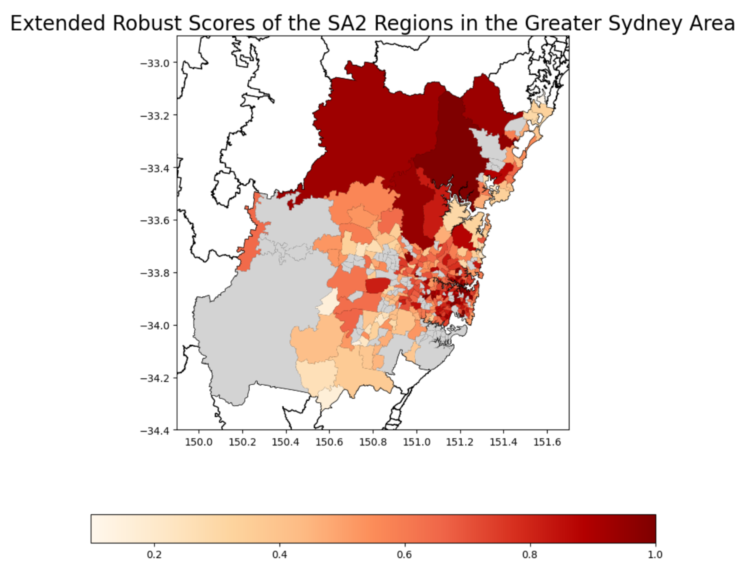
Discussions



The histogram of the scores distributions is shown on the right plot. For the z-score method, the distribution is almost homogeneous, with the low score end having slightly higher frequency. For the two robust z-score methods, however, the distributions appear to follow a normal curve, except there is a sudden rise at the highest score bin. This may represent the region within the CBDs which has the most resources.

Next, we plotted the choropleth of both robust and extended robust scores for each region to inspect whether the two display similar trends. The darker the red colour is, the higher the score is for a region. The regions whose score was not computed due to low population are filled with grey colour. Except there are more grey areas (missing value) in the extended version map, the colour pattern on both maps are similar, which means adding emissions and crime features does not change the overall trend of the scores.





Unsurprisingly, the SA2 regions closer to the CBD of Sydney had higher scores on average due to being the most populated regions of Sydney. However, there were also several rural regions west of Gosford such as Bilpin, Dural and Calga that unexpectedly had very high scores, given that their rural nature would theoretically result in them having a low number of bus stops and businesses compared to other regions, however upon further inspection this is in fact a product of the low population density of these regions where the low population amplifies scores of the per capita metrics. Apart from that, we see wealthier commercial suburbs such as Chatswood and Surry Hills (high number of polling centres, retail and health services) high on the list as well as more remote locations such as Katoomba-Leura which has a surprising amount of bus stops and Penrith that has a high amount of bus stops and polling areas. We also see suburbs that have a poor amount of bus stops and polls but that are generally considered good suburbs to live in such as Dover Heights, Cromer, Dee Why and North Curl Curl are penalised greatly and end up near the bottom of the list. (Please refer to the appendix for the complete list of the score for each region.)

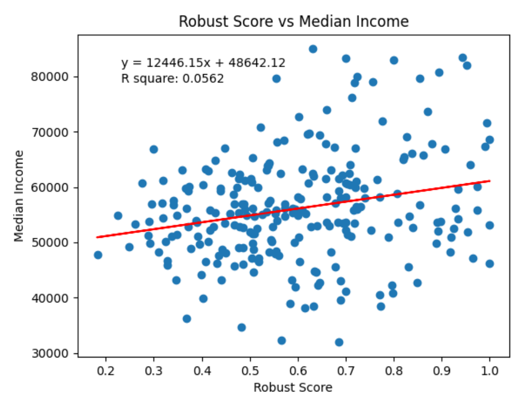
Limitations

Overall we see the number of bus stops and polling stations being disproportionately weighted and hence being biassed against many residential suburbs that generally possess car centric infrastructure and also lack a large number of businesses or schools to offset those poor z-scores. Due to this we see regions centred around the CBD as well as the more remote, yet frequently visited regions of Sydney of the Blue Mountains and Penrith being significantly favoured in this scoring system. These two factors being prioritised is also problematic considering that the number of bus stops alone gives an incomplete picture of public transport accessibility as train and ferry lines are not considered and given that federal elections occur rather infrequently and as such are likely a minor consideration for many when choosing a region to live in.

IV. Correlation Analysis

Significance Test on the Correlation Coefficient

In the scatter plot depicting the relationship between score and median income, it is apparent that the data points deviate considerably from the trendline, indicating a weak correlation between the variables. Nevertheless, there seems to be a discernible trend suggesting that higher scores tend to be associated with higher median incomes.



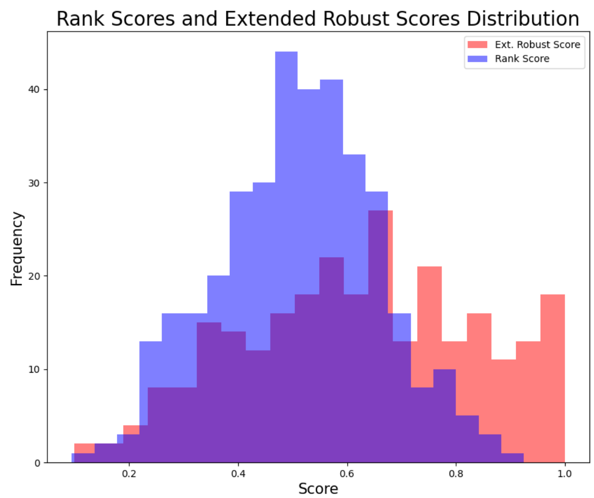
To verify this observation, we conducted a statistical analysis to evaluate the correlation between score and median income across the SA2 regions in Greater Sydney, employing Pearson's correlation coefficient (r). After eliminating missing values, the dataset consisted of 247 SA2 regions.

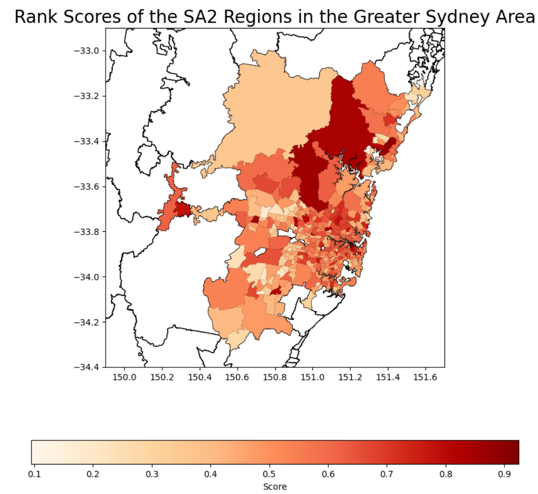
The calculated Pearson correlation coefficient was found to be r = 0.237. To ascertain if there is a positive correlation between score and median income, a t-test was performed on the correlation coefficient. The null hypothesis assumed no correlation (r = 0), while the alternative hypothesis posited a positive correlation (r > 0). With a sample size of n = 247 and a chosen significance level of $\alpha = 0.05$, the test statistic was calculated as $t_0 = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = 3.818$.

The resulting p-value was determined to be $P(t_{n-2} \geq t_0) = 8.519 \times 10^{-5}$, which is much lower than the chosen significance level (α). This suggests that, under the null hypothesis, the probability of observing a test statistic as extreme as or more extreme than the obtained result is 8.519×10^{-5}. Consequently, at the 0.05 significance level, we reject the null hypothesis and conclude that there exists a weak correlation between our score and median income within the SA2 regions of Greater Sydney.

Overall, it was surprising to learn that the correlation between the score and the median income was quite low. And it appears that income is not a good predictor of how 'well-resourced' a region is.

V. Additional Analysis

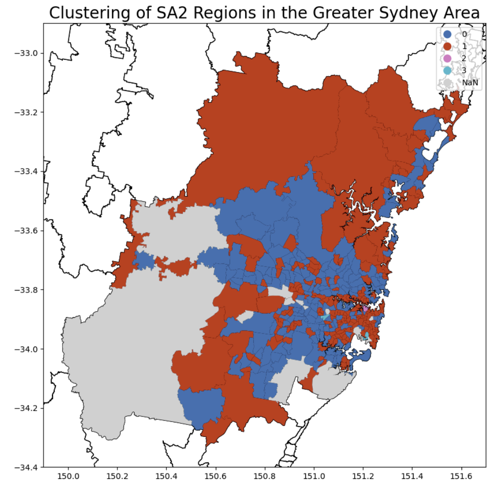




Overall, we found that our ranked algorithm produced a distribution that more closely resembled a normal distribution, or bell curve, than the other algorithms we tested. This means that the results were more evenly distributed, with fewer extreme values as we would expect assuming that the regions follow the central limit theorem. Hence this ranked scoring algorithm more closely reflects reality and makes the results more reliable and easier to interpret. We found a low correlation between our new algorithm vs the old ranking system, which hopefully means that we managed to correct some of the weighting flaws that the old scoring system had. We still encounter the issue of having nearly unpopulated rural regions as well as Penrith and the Blue Mountains region being given very high scores while the more isolated residential suburbs still rank lowly. So it does not appear that scoring by ranks alleviates any of the biases we first encountered.

To get a better understanding of the overall context of the SA2 regions and to identify any hidden relationships and trends between the regions, we applied a clustering algorithm to the majority of the data that we had gathered. We selected the K-means algorithm for its simplicity, scalability and widespread use. As there was a significant amount of SA2 regions that had missing emissions and median income data, we decided to exclude those variables from this analysis. We then used principal component analysis to reduce the amount of variable dimensions to three and make it more interpretable. Fully understanding what each of these dimensions represent is a nuanced topic, but a rough intuition is that the first component selects for the more industrial areas that have high school and business density with low population and accessibility, the second selects for commercial areas that have high amounts of businesses, accessibility and polling centres while also have higher populations and the third component generally prioritises residential areas that have high amounts of schools, accessibility of public transport and overall population.

Ultimately, we separated the regions into 4 clusters; the vast majority of regions were classified as blue or red, the two regions of Banksmeadow and Chullora were classified as green while the north part of main Sydney (Wynyard, Circular Quay, Millers Point) was the only region classified as green. The other regions coloured grey are unpopulated regions that we considered irrelevant due to being unlivable areas. The two green regions represented industrial areas that barely passed the minimum population of 100 and due to that yielded extremely high scores for metrics that relied on per capita calculations. This highlights a major flaw in our scoring calculations where we failed to account for the impact that low population regions would have on our metrics. We then have the outlier of the North tip of Sydney which is the biggest commercial area by far as represented by its high values in commercialness and industrialness. The rest of the regions are more tightly clustered but are divided mainly by their population with blue regions generally having higher populations, accessibility and less schooling. These regions are generally further away from the CBD, while the red regions of central and west Sydney as well as some far west and north regions tend to be lower populated but with higher business density. Unfortunately, there is some overall between these two main regions which makes it difficult to extrapolate further trends. This could perhaps be improved in the future by integrating more datasets that include data about recreational activities and areas available to further distinguish between the more commercial and more residential suburbs.



Appendix

Robust Scores

SA2 Name	Robust Score	SA2 Name	Robust Score
Sydney (North) - Millers Point	1	Blaxland - Warrimoo - Lapstone	0.881
Banksmeadow	1	Normanhurst - Thornleigh - Westleigh	0.879
Dural - Kenthurst - Wisemans Ferry	1	Strathfield - West	0.869
Sydney (South) - Haymarket	0.999	Umina - Booker Bay - Patonga	0.858
Chatswood - East	0.997	Bass Hill - Georges Hall	0.85
Chullora	0.996	Asquith - Mount Colah	0.85
Calga - Kulnura	0.995	St Ives	0.848
Baulkham Hills (West) - Bella Vista	0.995	St Leonards - Naremburn	0.844
Parramatta - North	0.994	Turramurra	0.834
Campbelltown - Woodbine	0.992	Mona Vale - Warriewood (North)	0.826
Penrith	0.986	Guildford - South Granville	0.826
Darlinghurst	0.982	Ermington - Rydalmere	0.821
Gosford - Springfield	0.978	Tuggerah - Kangy Angy	0.811
Bondi Junction - Waverly	0.969	Ryde - North	0.809
Surry Hills	0.968	Newtown (NSW)	0.803
Katoomba - Leura	0.966	Pennant Hills - Cheltenham	0.802
Erina - Green Point	0.961	Rose Bay - Vaucluse - Watsons Bay	0.797
Lindfield - Roseville	0.957	Frenchs Forest - Oxford Falls	0.792
Freshwater - Brookvale	0.955	Gordon - Killara	0.786
Springwood - Winmalee	0.952	Rooty Hill - Minchinbury	0.786
Pymble	0.947	West Pennant Hills	0.783
Double Bay - Darling Point	0.942	Paddington - Moore Park	0.779
North Sydney - Lavender Bay	0.936	Balgowlah - Clontarf - Seaforth	0.776
Miranda - Yowie Bay	0.926	Pitt Town - McGraths Hill	0.771
Kurrajong Heights - Ebenezer	0.926	Randwick - South	0.77
Baulkham Hills - East	0.921	Condell Park	0.767
Eastwood	0.918	Emu Plains - Leonay	0.767
Hurstville - Central	0.904	Menai - Lucas Heights - Woronora	0.764
Blacktown (East) - Kings Park	0.903	Concord - Mortlake - Cabarita	0.762
Wahroonga (East) - Warrawee	0.9	South Hurstville - Blakehurst	0.762
Lalor Park - Kings Langley	0.9	Chittaway Bay - Tumbi Umbi	0.756
Jilliby - Yarramalong	0.893	Castle Hill - Central	0.756
Greenwich - Riverview	0.893	Terrigal - North Avoca	0.75
Ashfield - South	0.75	Erskineville - Alexandria	0.621
Gorokan - Kanwal - Charmhaven	0.748	Mosman - North	0.62

Lidcombe	0.748	Ingleburn	0.617
Warnervale - Wadalba	0.746	Bradbury - Wedderburn	0.613
Bankstown - South	0.744	Wamberal - Forresters Beach	0.613
Drummoyne - Rodd Point	0.74	Neutral Bay - Kirribilli	0.612
Cabramatta - Lansvale	0.74	North Parramatta	0.611
Chipping Norton - Moorebank	0.728	Smithfield - Wetherill Park	0.61
Castle Cove - Northbridge	0.722	Bellevue Hill	0.606
Camden - Ellis Lane	0.715	Crows Nest - Waverton	0.605
Kingswood - Werrington	0.715	Toongabbie - Constitution Hill	0.602
Manly - Fairlight	0.714	Bilpin - Colo - St Albans	0.601
Mosman - South	0.704	Holsworthy - Wattle Grove	0.599
Haberfield - Summer Hill	0.703	Northmead	0.598
Kingsgrove (South) - Bardwell Park	0.703	Claymore - Eagle Vale - Raby	0.597
Cronulla - Kurnell - Bundeena	0.702	St Marys - North St Marys	0.597
Merrylands - Holroyd	0.701	Engadine	0.594
Glebe - Forest Lodge	0.698	Sylvania - Taren Point	0.593
Silverwater - Newington	0.697	Rouse Hill - Beaumont Hills	0.587
Macquarie Park - Marsfield	0.697	Castle Hill - North	0.583
Bateau Bay - Killarney Vale	0.689	Leichhardt	0.579
Potts Point - Woolloomooloo	0.687	Blackheath - Megalong Valley	0.576
Oyster Bay - Como - Jannali	0.686	Rockdale - Banksia	0.576
Hornsby - West	0.679	Richmond - Clarendon	0.572
Woy Woy - Blackwall	0.675	Kogarah	0.568
Avalon - Palm Beach	0.675	Windsor - Bligh Park	0.566
Strathfield - East	0.671	Mount Druitt - Whalan	0.566
Galston - Laughtondale	0.663	Fairfield - East	0.562
Hunters Hill - Woolwich	0.662	The Oaks - Oakdale	0.558
Lane Cove	0.659	Hornsby - East	0.555
Burwood (NSW)	0.658	Randwick - North	0.552
Terrey Hills - Duffys Forest	0.656	Rosemeadow - Glen Alpine	0.551
Peakhurst - Lugarno	0.653	Marrickville - South	0.549
Austral - Greendale	0.644	Sydenham - Tempe - St Peters	0.545
Sutherland - Kirrawee	0.639	Minto - St Andrews	0.543
Horsley Park - Kemps Creek	0.636	Leumeah - Minto Heights	0.542
Chester Hill - Sefton	0.631	Carlingford - West	0.536
North Ryde - East Ryde	0.631	West Ryde - Meadowbank	0.536
Liverpool - East	0.628	Granville - Clyde	0.532
Five Dock - Abbotsford	0.627	Cherrybrook	0.523
Woollahra	0.626	Coogee - Clovelly	0.521
Lilyfield - Rozelle	0.52	Croydon	0.415
Chatswood (West) - Lane Cove North	0.515	Monterey - Brighton-le-Sands - Kyeemagh	0.409

Campsie - South	0.513	Picton - Tahmoor - Buxton	0.406
Belmore - Belfield	0.512	Cranebrook - Castlereagh	0.402
Gladesville - Huntleys Point	0.511	Cremorne - Cammeray	0.399
Bankstown - North	0.51	Wentworthville - Westmead	0.399
Guildford West - Merrylands West	0.51	Kensington (NSW)	0.392
Greenacre - North	0.509	Wyoming	0.392
Camperdown - Darlington	0.508	Chippendale	0.391
Greystanes - South	0.503	Annandale (NSW)	0.388
Lethbridge Park - Tregear	0.502	Bondi - Tamarama - Bronte	0.388
Ourimbah - Fountaindale	0.499	North Rocks	0.386
Doonside - Woodcroft	0.499	Castle Hill - East	0.384
Revesby	0.496	Leppington - Catherine Field	0.379
Punchbowl	0.495	Willoughby	0.378
Blacktown (North) - Marayong	0.493	Roselands	0.377
Wyong	0.486	Manly Vale - Allambie Heights	0.376
The Entrance	0.479	Kincumber - Picketts Valley	0.376
Padstow	0.472	Prestons	0.368
Fairfield	0.469	Elderslie - Narellan	0.366
Earlwood	0.464	Bidwill - Hebersham - Emerton	0.363
Caringbah	0.463	Marrickville - North	0.362
Mulgoa - Luddenham - Orchard Hills	0.46	Ashcroft - Busby - Miller	0.362
Bonnyrigg Heights - Bonnyrigg	0.456	Riverwood	0.356
Yagoona - Birrong	0.454	Auburn - Central	0.354
Strathfield South	0.452	Dee Why - North	0.348
Petersham - Stanmore	0.445	Casula	0.346
Epping (NSW) - West	0.444	Ultimo	0.345
Sans Souci - Ramsgate	0.444	Box Head - MacMasters Beach	0.342
Cabramatta West - Mount Pritchard	0.439	Liverpool - West	0.337
Kellyville - East	0.437	Seven Hills - Prospect	0.337
Glenmore Park - Regentville	0.436	Maroubra - North	0.335
Riverstone	0.429	Putney	0.333
Gymea - Grays Point	0.427	Carlingford - East	0.329
Oatlands - Dundas Valley	0.426	Yarramundi - Londonderry	0.323
Pendle Hill - Girraween	0.425	Redfern	0.322
Jamisontown - South Penrith	0.425	Panania (South) - Picnic Point	0.322
Kogarah Bay - Carlton - Allawah	0.42	Rosebery - Beaconsfield	0.319
Balmain	0.417	Lawson - Hazelbrook - Linden	0.319
Berowra - Brooklyn - Cowan	0.416	Oatley - Hurstville Grove	0.316
Glenhaven	0.416	Quakers Hill	0.316
Bondi Beach - North Bondi	0.315	Epping (East) - North Epping	0.222
Box Hill - Nelson	0.314	Blacktown - West	0.219

Douglas Park - Appin	0.304	Malabar - La Perouse	0.216
Narwee - Beverly Hills	0.303	Hinchinbrook	0.213
Hurstville - North	0.3	Kellyville - West	0.21
Bossley Park - Abbotsbury	0.293	Mount Annan	0.209
Macquarie Fields	0.29	Winston Hills	0.208
Toukley - Norah Head	0.288	Cobbitty - Bringelly	0.207
Illawong - Alfords Point	0.288	Stanhope Gardens - Parklea	0.205
Canterbury (North) - Ashbury	0.286	Beacon Hill - Narraweena	0.205
Avoca Beach - Copacabana	0.284	Fairfield - West	0.204
Auburn - North	0.284	Kellyville Ridge - The Ponds	0.2
Matraville - Chifley	0.283	Warwick Farm	0.196
Greenacre - South	0.282	Eastlakes	0.195
Cambridge Park	0.278	Waterloo	0.193
Kingsgrove - North	0.275	Bexley - North	0.187
Maroubra - South	0.274	Wentworth Falls	0.187
Auburn - South	0.272	Mortdale - Oatley	0.182
Dulwich Hill - Lewisham	0.272	Campsie - North	0.18
Concord West - North Strathfield	0.272	Glenfield	0.179
Bayview - Elanora Heights	0.27	Loftus - Yarrawarrah	0.177
Croydon Park - Enfield	0.268	South Wentworthville	0.176
Glenwood	0.262	Budgewoi - Buff Point - Halekulani	0.173
Currans Hill	0.258	Homebush	0.172
Bexley - South	0.254	Green Valley	0.171
Arncliffe - Bardwell Valley	0.251	Pemulwuy - Greystanes (North)	0.164
St Clair	0.251	Caringbah South	0.164
Regents Park	0.248	Glendenning - Dean Park	0.164
Newport - Bilgola	0.247	Narrabeen - Wheeler Heights	0.164
Woolooware - Burraneer	0.246	Lurnea - Cartwright	0.162
Castle Hill - South	0.244	Point Clare - Koolewong	0.162
Penshurst	0.242	Schofields - East	0.162
Forestville - Killarney Heights	0.24	Kingsford	0.16
Lake Munmorah - Mannering Park	0.239	St Johns Park - Wakeley	0.157
Hassall Grove - Plumpton	0.237	Cecil Hills	0.157
Lakemba	0.236	Cromer	0.153
Belrose	0.235	Rhodes	0.152
Panania (North) - Milperra	0.225	Mascot	0.149
North Narrabeen - Warriewood (South)	0.224	Pyrmont	0.146
Dover Heights	0.222	Maroubra - West	0.145
Castle Hill - West	0.143	Warragamba - Silverdale	0.067
Colyton - Oxley Park	0.139	Canterbury - South	0.067

Erskine Park	0.136	Parramatta - South	0.066
Greenfield Park - Prairiewood	0.136	Jordan Springs - Llandilo	0.065
Pagewood - Hillsdale - Daceyville	0.136	Marsden Park - Shanes Park	0.061
West Hoxton - Middleton Grange	0.133	Edmondson Park	0.061
Denistone	0.132	Kariong	0.054
Wentworth Point - Sydney Olympic Park	0.127	Schofields (West) - Colebee	0.052
Rosehill - Harris Park	0.125	Summerland Point - Gwandalan	0.041
Gledswood Hills - Gregory Hills	0.124	Spring Farm	0.036
South Coogee	0.124	Prospect Reservoir	NA
Ryde - South	0.121	Port Botany Industrial	NA
Berala	0.117	Sydney Airport	NA
Artarmon	0.116	Centennial Park	NA
Heathcote - Waterfall	0.114	Holsworthy Military Area	NA
Collaroy - Collaroy Plateau	0.113	Blue Mountains - North	NA
Lilli Pilli - Port Hacking - Dolans Bay	0.109	Blue Mountains - South	NA
Zetland	0.109	Rookwood Cemetery	NA
Toongabbie - West	0.109	Smithfield Industrial	NA
Woronora Heights	0.106	Yennora Industrial	NA
Blacktown - South	0.105	Badgerys Creek	NA
Niagara Park - Lisarow	0.101	Wetherill Park Industrial	NA
Hoxton Park - Carnes Hill - Horningsea Park	0.097	Royal National Park	NA
North Kellyville	0.094		
Saratoga - Davistown	0.094		
Harrington Park	0.094		
Dee Why (South) - North Curl Curl	0.094		
Wahroonga (West) - Waitara	0.092		
Blue Haven - San Remo	0.089		
Botany	0.089		
Ashfield - North	0.087		
Edensor Park	0.085		
Acacia Gardens	0.081		
Wolli Creek	0.08		
Wiley Park	0.078		
Bargo	0.076		
Oran Park	0.076		
Narara	0.073		
Denham Court - Bardia	0.069		

V. Bibliography

1. Australian Bureau of Statistics. (Jul2021-Jun2026). *Digital boundary files*. ABS. https://www.abs.gov.au/statistics/standards/australian-statistical-geography-standard-asgs-edition-3/jul2021-jun2026/access-and-downloads/digital-boundary-files.
2. Australian Bureau of Statistics. (Jul2018-Jun2022). *Counts of Australian Businesses, including Entries and Exits*. ABS. https://www.abs.gov.au/statistics/economy/business-indicators/counts-australian-businesses-including-entries-and-exits/latest-release.
3. Transport for NSW. (2022, November 1). *Timetables Complete GTFS.* https://opendata.transport.nsw.gov.au/dataset/timetables-complete-gtfs
4. NSW Department of Education. (2023, May 18). *School intake zones (catchment areas) for NSW government schools.* *https://data.cese.nsw.gov.au/data/dataset/school-intake-zones-catchment-areas-for-nsw-government-schools*
5. Australian Electoral Commission. (2023, April 27). AEC - Federal Election - Polling Places (Point) 2019. https://data.aurin.org.au/dataset/au-govt-aec-aec-federal-election-polling-places-2019-na
6. Australian Bureau of Statistics. (2021). *Regional population by age and sex*. ABS. https://www.abs.gov.au/statistics/people/population/regional-population-age-and-sex/2021.
7. Australian Bureau of Statistics. (2015-16-to-2019-20). *Personal Income in Australia: Table 1 - Total income, earners and summary statistics by geography, 2015-16 to 2019-20*. ABS. https://www.abs.gov.au/statistics/labour/earnings-and-working-conditions/personal-income-australia/latest-release.
8. https://datasets.seed.nsw.gov.au/dataset/nsw-regional-and-local-greenhouse-gas-emissions-2016-2019/resource/fbbafe30-e254-4d9a-816b-7bcba989bf97
9. NSW Bureau of Crime Statistics and Research (2023, May 19). Crime Hotspots 2022. https://www.bocsar.nsw.gov.au/Pages/bocsar_datasets/Spatial.aspx