1 Dataset Description

The following data sources were used in this analysis:

- **SA2 Regions:** Geographic and metadata for all SA2 regions, used to join all other datasets spatially. Derived from the ABS shapefile. ABS SA2 Reference Page.
- **Businesses.csv:** Business counts by income category and SA2, sourced from the Australian Bureau of Statistics. ABS Business Counts.
- **Stops.txt:** Public transport stop data from Transport for NSW. GTFS Complete Timetables.
- **SchoolCatchments.zip:** School zone catchments from the NSW Department of Education. NSW School Catchment Boundaries.
- **POIs:** Points of Interest data collected using the NSW POI API and filtered by selected categories (Task 2). NSW POI API.
- **Population and Income:** Demographic counts and income data for each SA2, including total population, age bands, and median income. ABS Population Data.

1.1 Definition of a Well-Resourced Region

In our analysis, we define a well-resourced SA2 or SA4 region as one that meets the essential needs for daily life like access to education, transportation, commercial services, and essential amenities. This definition influenced how we cleaned our data. For example, when looking at the business dataset, we decided to exclude sectors like agriculture and mining because they're generally resource or land-heavy and don't truly show the availability of local services. We also filtered out hydrography-related entries, such as lakes and rivers, in the POI dataset. While these features are geographically significant, they don't serve as practical need for residents. This targeted approach helps ensure that our composite scoring system accurately represents the services that enhance everyday quality of life (Badland et al., 2014). Also, when calculating the z-score for school, we refined the analysis by including only the population aged 5–19 years. This age range captures the core members in the society that would be benefit from having primary or secondary education. By doing this, we provide a more accurate reflection of whether an area is adequately resourced for schooling relative to the needs of its residents.

2 Database Description

The PostgreSQL database was structured using a dedicated schema `sydney_analysis`, created via SQLAlchemy and PostGIS. All relevant datasets were imported as relational tables, linked by SA2 codes. A spatial reference system (SRID 4283) was used for geospatial analysis.

- **sa2_regions:** Core table with geometries and metadata for each SA2.
- **businesses, stops, schools, poi, population, income:** Thematic tables referencing SA2 and containing variables used in scoring.

A relational database management system (RDBMS) organizes data into tables, where each table represents an entity (e.g., businesses, schools, SA2 regions). Each table includes a primary key and a unique identifier for each row (e.g., `sa2_code21` in the sa2_regions table). To model relationships between tables, foreign keys are used — for example, the businesses, schools, and stops tables each include `sa2_code21` as a foreign key that references the sa2_regions table. This relational structure ensures referential integrity, allowing efficient data joins and maintaining consistency across the database schema.

For indexes, in our project, we created several spatial and attribute indexes.

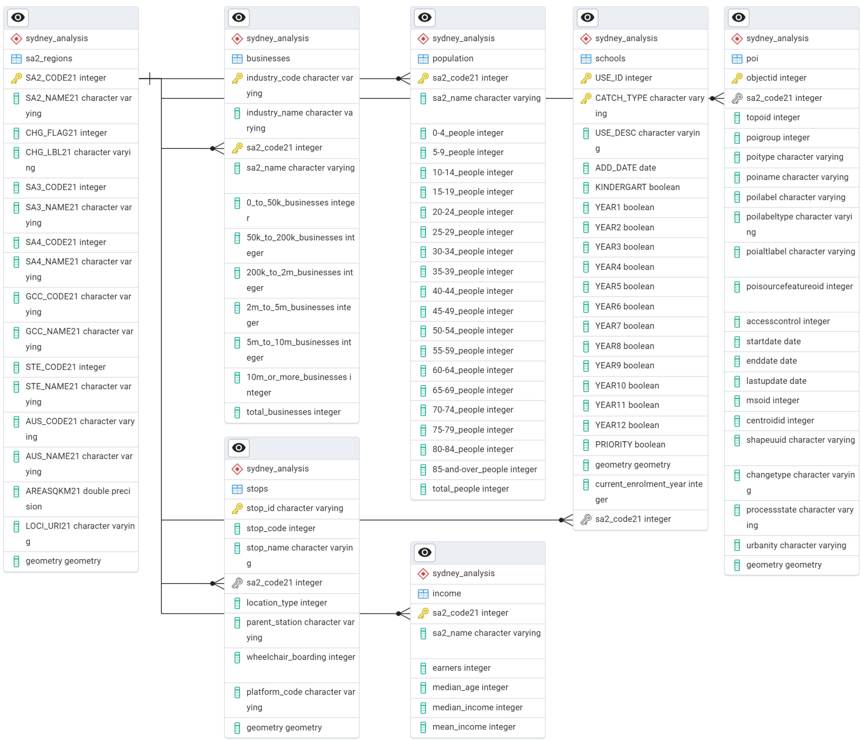
- **Spatial Index:** A spatial `GIST` index `(sa2_regions_geom_idx)` was created on the `geometry` column of the `sa2_regions` table. This index accelerates spatial operations such as `ST_Within` and `ST_Intersection`, which are frequently used to assign SA2 codes based on geographic boundaries.

- **Attribute Index (SA4 Filter):** A B-tree index `(sa2_regions_idx)` was created on the `SA4_CODE21` column to allow efficient filtering of SA2s by SA4 region.

- **Attribute Indexes on SA2 Codes:** To support fast lookups and joins, B-tree indexes were created on the `sa2_code21` column across all core tables: `businesses`, `stops`, `schools`, `population`, `income`, and `poi`. These indexes enhance performance for queries that involve grouping or filtering based on SA2 codes, which are central to our spatial and demographic analysis.

Together, these indexes significantly reduced the execution time of spatial joins and aggregations, especially when processing large-scale datasets.

Also, PostGIS extensions enabled geometry-based queries and joins. The following diagram shows the full schema with relationships:



3 Result Analysis

3.1 Comparison Between SA4s

1. **Rationale for the Formula Computation**

   Each SA2 score was calculated by finding the z-scores with four key indicators: business density (per 1,000 residents), number of schools(only 5-19 years old), public transport stops, and POIS (Points of Interest). They were selected based on their capturing accessibility and service availability within a region. The z-scores standardise. The final score is passed through a sigmoid function to bound the values and reduce outlier effects. However, for the business dataset, we deleted the column for Agriculture, Mining, Manufacturing, Construction, wholesale trade, Professional services and Art, recreation centre. These businesses do not directly relate to the basic living needs of a citizen, while the remaining businesses like Accommodation, Transport, etc are essential to daily life. For POIS, we have only included POIGroups 1-5, which are Community, Education, Recreation, Transportation and Utility. These are again, essential to daily life, while Hydrography Landform are not that crucial when compared.

2. **Implications of Extending the Formula**

   The scoring formula could incorporate additional datasets, such as healthcare, crime rates or internet access. This would broaden the analysis and provide a more comprehend view of regional livability.

3. **Summary of Overall Distribution**

   The composite accessibility scores across 98 SA2 regions range from a minimum of 0.0018 to a maximum of nearly 1. This wide range suggests significant variability in the concentration and distribution of services, infrastructure, and amenities across Greater Sydney.

   - **Minimum score:** 0.0018 — indicating a region with very limited access to the selected facilities.
   - **Maximum score:** 0.9999 — denoting an exceptionally well-served area.
   - **15.3% (15 out of 98)** of SA2s fall within the moderate range (0.4 to 0.7), implying that most regions skew toward either high or low accessibility extremes rather than the middle.

   This uneven distribution highlights spatial inequality, where a small number of SA2s benefit from a high concentration of essential services, while others remain under-resourced.

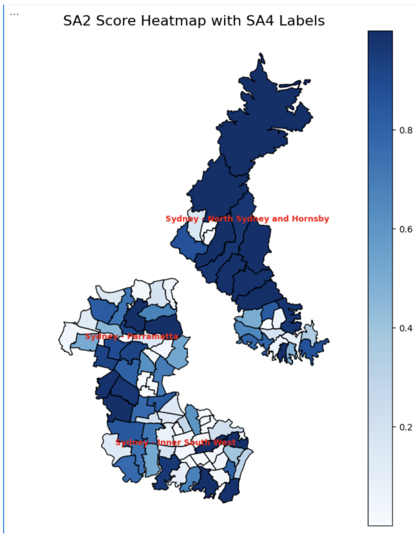
4. **Score Differences Across SA4 Zones**

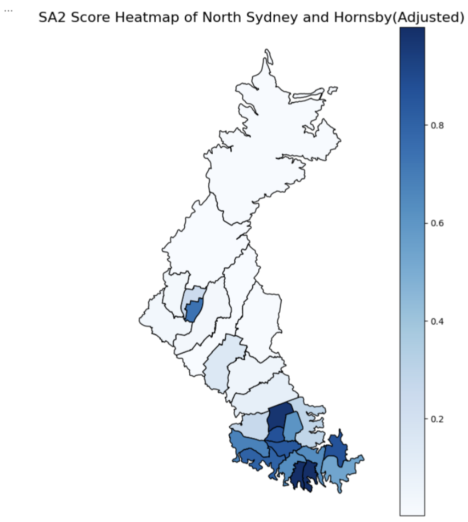
   - **North Sydney and Hornsby (SA4: 121)** initially appears to have consistently high scores. However, when adjusted for area, the central and northern SA2s within this region show significantly lower scores. More detailed explanation below in "Trends and Regions Of Interest" section.
   - **Inner South West (SA4: 119)** shows a more variable pattern—Bankstown and Canterbury areas perform well, but other regions lag.
   - **Parramatta (SA4: 125)** has generally lower scores, with a few standout areas likely benefiting from recent infrastructure investment.

5. **Trends and Regions of Interest** Click here to open the interactive map

   High-scoring SA2s are generally concentrated in urban regions with high resource density and strong public transport infrastructure. Areas within Parramatta and the Inner South West display composite scores above 0.8, driven by balanced concentrations of businesses, schools, and POIs, making them well-resourced and accessible.

   In contrast, outer metropolitan SA2s, such as parts of the southern and western boarder, tend to score lower due to limited infrastructure and reduced access to services, reflecting their ongoing urban development or transitional planning stage(unknown, 2022).



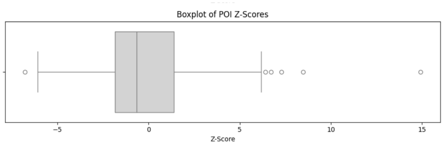


Additional Insight – Discrepancy in North Sydney and Hornsby Region

A notable finding arises in the northern segment of the *North Sydney and Hornsby* SA4, especially in SA2s such as *Hornsby East*, *Wahroonga*, and *Berowra*. These regions show unexpectedly high Z scores in the main heatmap(Left Figure), primarily because of the presence of many transport-related POIs. However, these SA2s are large geographical areas and have low population density and don't really have much urban infrastructure. To explore this further, an additional map (Right Figure) presents scores adjusted by area, which reveals that these regions actually have among the lowest service densities in the selected SA4s.

This difference suggests that the high scores in the original visualization are reflecting how spatial scale—larger SA2s can accumulate more POIs simply due to their size, even if those services are spread thinly. While area-normalized scores are not used in the final composite scoring to maintain methodological consistency, this comparative view highlights the importance of spatial context. It reinforces that SA2s like *Berowra* may appear "well-resourced" on paper but remain underserved in practice.

6. **Outlier Explanation**



According to the Box plot generated for POI, there are a few outliers.

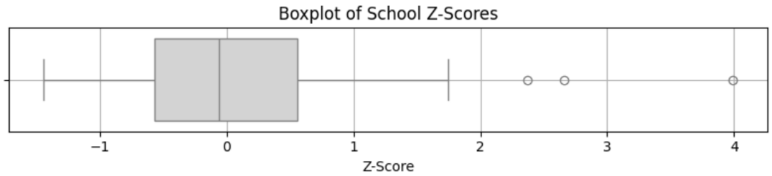
Right-side outliers (z-scores larger than 7, one as high as ∼15):

- These SA2s have extremely high concentrations of POIs.
- Likely inner-city or commercial hubs
- This suggests a very dense clustering of amenities, such as shopping centres, dining areas, and public institutions, which significantly skews the overall distribution.

Left-side outlier (z-score ≈ -6):

- A region with very few POIs relative to the national average.
- Likely outer urban or semi-rural border areas

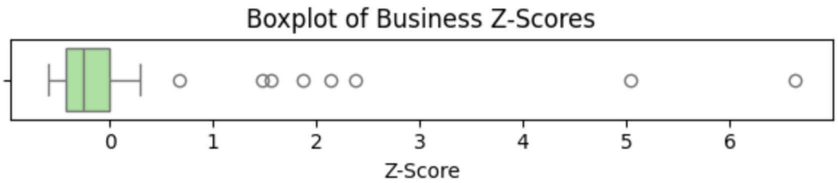
- Such SA2s may lack urban infrastructure or are in early development stages.
- highlight underserved regions that could benefit from targeted urban planning or investment.



According to the Box Plot generated for Schools,

Right-side outliers (z-scores between 2.3 and 4):

- A small number of SA2s have significantly above-average school counts.
- likely urban centres or education hubs that serve broader catchment areas, such as *Parramatta*, *Bankstown*, or *Chatswood*.
- Implicate that these regions may be well-established with dense educational facilities, potentially skewing the metric if not normalized.



According to the Box Plot generated for Business,

Right-side outliers (Z-score > 1.5 up to ∼6.8):

- These SA2 regions shows unusually high business density per 1,000 residents.
- Likely commercial or mixed-use urban centres such as *Parramatta*, or *Chatswood*.
- These areas typically have dense clusters of retail, hospitality, and office-based services.
- **Implication:** While normalization via z-scores helps, further transformation (e.g. sigmoid scaling) ensures these outliers do not dominate.

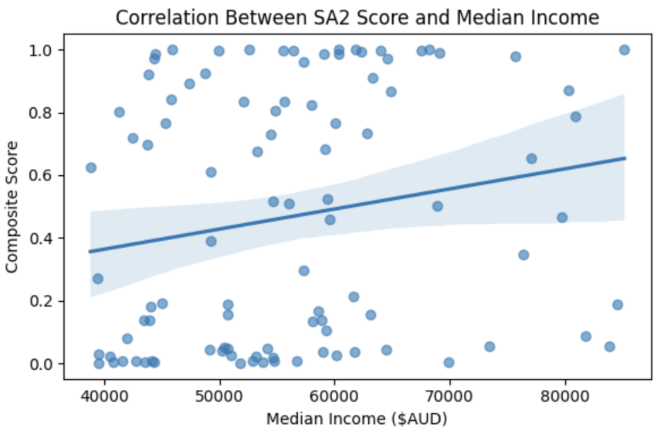
4 Correlation Analysis

To evaluate the relationship between SA2 composite scores and economic prosperity, a Pearson correlation test was performed using median income data for each SA2 region.

- **Pearson correlation coefficient:** 0.186
- **P-value:** 0.0667

This result indicates a weak positive correlation between Z scores and median income. However, the p-value exceeds the standard 0.05 threshold, suggesting the relationship is not statistically significant. In other words, while there is a slight tendency for better-resourced SA2s to correspond with higher income levels, the evidence is not strong enough to confirm this trend with confidence.

Scatterplot Interpretation



This scatterplot illustrates the relationship between the computed SA2 composite scores and the corresponding median incomes. A weak upward trend is visible, but the data points are widely scattered.

- **Distribution:** Composite scores are distributed sparsely across all income ranges, with several high-scoring SA2s existing at both low and high income levels. This suggests that factors beyond income, such as planning, infrastructure investment, or population density—also contribute to accessibility.
- **Outliers:** Some low-income SA2s still achieve high composite scores, which may reflect areas with focused public investment or high service density despite economic disadvantage. Likewise, a few high-income SA2s have moderate scores, possibly due to lower service density or geographic spread.

The score captures accessibility and resource availability that are not solely driven by economy. While this enhances the score's usefulness for identifying regions with strong service or infrastructure presence, it may also lead to misleading interpretations if used as a direct reference for wealth or quality of life. High scores in low-income areas may reflect government planning priorities or population density, not affluence. Conversely, some wealthy suburbs may score lower due to zoning or sparse amenity distribution. This suggests that while the composite score is a valuable spatial indicator for urban planning, it should be interpreted alongside socioeconomic indicators rather than in isolation.

References

Badland, H., Whitzman, C., Lowe, M., Davern, M., Aye, L., Butterworth, I., Hes, D., and Giles-Corti, B. (2014). What makes a city liveable? *Cities*, 38:S51–S58.

unknown, A. (2022). Access to urban infrastructure: Inequality and spatial patterns. *Journal of Urban Studies or similar*. Accessed May 2025.