Greater Sydney Analysis Report

AUTHORS:- Ayush Arora (540906543), Lakshya Sakhuja (540863213)

1. Dataset Description

This report integrates several key datasets for a comprehensive analysis of resource availability across SA2 regions in Greater Sydney:

- **SA2 Boundaries (ABS):** Provided as a shapefile, this dataset outlines the official geographic boundaries of each SA2 statistical region within Greater Sydney. It forms the spatial foundation for all geospatial joins and analyses, using EPSG:4283 as the spatial reference system. We did not perform much data cleaning on this file as this file was taken from the government website and already must have undergone cleaning before being published.
- **Population Data (ABS):** A CSV file containing age-segmented population counts for each SA2. It is crucial for computing per capita measures, especially for youth-focused calculations (ages 0–19) tied to school catchments.
- **Income Data (ABS):** Contains the median weekly income per SA2. This dataset supports our correlation analysis to explore how well-resourced scores align with economic wellbeing.
- **Business Data (ABS):** Compiled from official business counts by industry type and SA2 region. It includes turnover ranges and allows the selective focus on sectors (e.g., retail, healthcare) during score computation.
- **Public Transport Stops (GTFS):** Extracted from Transport for NSW's GTFS feed, this dataset includes the location of all train and bus stops. It supports computation of stop density per square kilometer.
- **School Catchments (NSW Department of Education):** Delivered as multiple shapefiles representing primary, secondary, and future intake zones. This spatial data is intersected with population distributions to assess educational access per young person.
- **Points of Interest (NSW Spatial API):** Dynamically collected using bounding boxes of each SA2 from the NSW Points of Interest API. Returned features include locations of essential and lifestyle amenities like health centres, retail zones, and entertainment venues.

2. Database Description

The database was implemented in PostgreSQL with PostGIS enabled to support spatial operations. All datasets were ingested into a consistent and normalized schema, and linked through well-defined primary and foreign key constraints. Data types were carefully selected for efficiency and consistency, and indexing was used to optimize spatial joins and querying.

Schema Overview

- **sa2_regions**: This is the central table containing SA2 boundaries and names. The primary key is 'sa2_code' which serves as the reference point for all other data-sets.
- **population_data**: This table contains SA2-level population breakdowns by age group. 'sa2_code' is the primary key.
- **income_data**: Stores median income values per SA2 which are keyed by 'sa2_code'.
- **business_data**: Counts the number of businesses per SA2 by industry category. 'id' is the primary key and 'sa2_code' is the foreign key which links to 'sa2_regions'.
- **stops_data**: Contains public transport stops with spatial coordinates. Primary key is 'stop_id' and 'sa2_code' acts as a foreign key.

- **school_catchments_primary, school_catchments_secondary, school-catchments_future**: These are shapefiles which represent intake zones. Each uses ‘use_id’ as a primary key and was spatially linked to population and SA2 data.
- **selected_sa4_pois**: This table stores points of interest retrieved via the NSW POI API. Primary key is ‘id’ and a foreign key to ‘sa2_regions’.
- **well_resourced_scores**: Stores the final score computed per SA2. ‘id’ is the primary key and a foreign key ‘sa2_code’ links back to the ‘sa2_regions’.
- **pt_accessibility_scores**: This table stores the accessibility metrics such as stop_density_per_sqkm, z_score which are linked via ‘sa2_code’
- **sa2_greater_sydney:** This table contains spatial and identifier data for SA2 regions in Greater Sydney, used as the base for geographic joins and visualizations.

The entity relationship diagram provided above helps to visually summarise the relationships between tables with primary keys marked and foreign keys indicated via arrows. All geometry fields were spatially indexed enabling efficient spatial joins and distance-based filtering operations

3.1. Result Analysis

3.1.1. Formula Rationale

To assess how well-resourced each SA2 region is, we developed a composite score combining four key indicators: business density, public transport stops, school catchment coverage, and points of interest (POIs). Each component was standardised using z-scores and summed before applying a sigmoid transformation to scale results between 0 and 1.

This approach ensures comparability across variables with different units and reduces the effect of outliers. The resulting score provides an interpretable index of public resource availability for each SA2.

Each component represents a specific aspect of resource availability:

- **Businesses per 1000 residents**: A proxy for economic opportunities and services.
- **Public transport stop density**: Reflects mobility and commuting access.
- **School catchments per 1000 youth (ages 0–19)**: Captures educational access.
- **Points of Interest (POI)**: Measures proximity to lifestyle, retail, and health facilities.

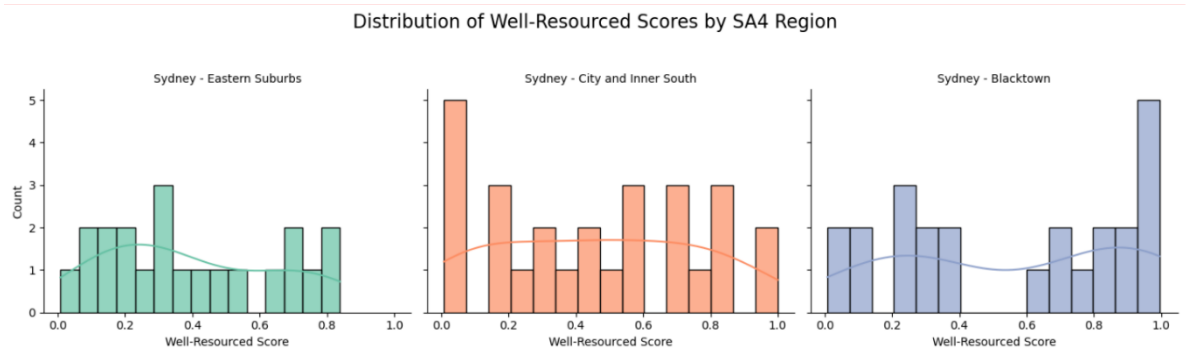
3.1.2. Impact of Different Components on Overall Score

The well-resourced score is a composite indicator that brings together several key public-facing dimensions, including access to health services, education, public infrastructure, and transport. Each of these components contributes in different ways across various regions. For instance, in some SA2 areas, strong access to public schools or close proximity to hospitals can significantly increase the score. On the other hand, certain regions with relatively high income levels may receive lower scores due to a lack of public facilities, which could reflect a greater reliance on private services.

This variation shows that no single factor determines the overall score. Instead, it is the combination and availability of multiple types of public resources that collectively define how well-resourced a particular area is. It also helps explain why some lower-income regions can still perform well in the index—because targeted public investment in infrastructure and essential services can meaningfully improve access, even in areas facing economic disadvantage (Dodson & Sipe, 2008).

.

3.1.3. Overall Distribution



- **Sydney - Eastern Suburbs**: The scores are slightly skewed toward higher values, suggesting a more consistent distribution of well-resourced areas across SA2 regions.
- **Sydney - City and Inner South**: This area displays a relatively even spread across the range of scores, but with some concentration in both low and high segments, reflecting mixed levels of resourcing.
- **Sydney - Blacktown**: This region shows a bimodal distribution with clear peaks in both high and low score ranges, indicating a divide in well-resourced access within the SA2s in this area.

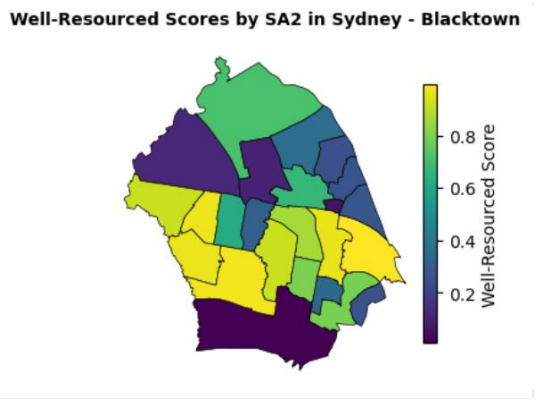
These visualizations justify the diversity within each SA4 region and highlights the disparities in resource distribution at a more local SA2 level. The skew is reflective of structural inequalities in urban planning:

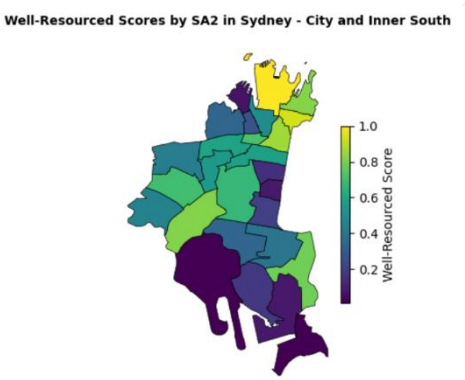
- High-scoring SA2s typically cluster around Sydney's CBD or Inner West, where infrastructure and services are concentrated.
- Lower-scoring regions tend to lie on the suburban fringe, particularly in parts of Blacktown where access to POIs and transport is limited.

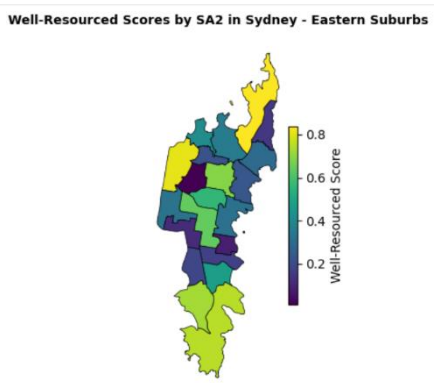
This visual evidence suggests that at the SA2 level, better planning is required through proper service distribution and development. Visualizing this distribution with a histogram confirms this imbalance. This insight justifies the use of normalisation and scaling in the scoring method and supports further discussion on policy-driven interventions to uplift lower-scoring SA2s (SGS Economics & Planning, 2019).

3.1.4. Comparison of Scores for Selected SA4 regions

We compared well-resourced scores across three SA4 regions: Blacktown, City and Inner South, and Eastern Suburbs, using maps that visualise SA2-level variation.







The well-resourced scores in Sydney - Blacktown show a stark contrast, with high-scoring SA2s in the north and east, and significantly under-resourced areas concentrated in the southern and western parts.

In City and Inner South, northern SA2s near the CBD score highest (0.9–1.0), while southern areas like Waterloo and Redfern show significantly lower scores, pointing to a stark urban contrast within close proximity.

The Eastern Suburbs present a more mixed pattern, with higher scores in Bondi and Maroubra but lower values in central pockets. This suggests resource concentration in some zones despite overall affluence.

Across all three SA4 zones, the maps reinforce the need for local SA2-level planning rather than broad SA4-level assumptions. These geographic disparities are critical for targeted policy interventions aimed at improving equity in public service delivery (NSW Department of Planning and Environment, 2022)

3.1.5. Trends and Noteworthy Patterns

Several noteworthy patterns emerge from the spatial analysis of well-resourced scores:

- **Urban Core Advantage**: SA2 regions near central business hubs consistently show higher scores. For instance, **Sydney CBD**, **Surry Hills**, and **Pyrmont–Ultimo** in the City and Inner South SA4 stand out with scores above 0.8. This aligns with expectations—urban cores tend to be better serviced due to historical investment, higher population density, and greater political and economic focus.
- **Bimodal Clustering**: Particularly in Blacktown, a **bimodal trend** is visible—some areas are very well resourced while others are significantly underserved, with few in the middle. This may reflect uneven urban development or prioritization of certain growth corridors.
- **Peripheral Underservice**: In contrast, **southern Blacktown SA2s**, such as **Doonside** and **Plumpton**, cluster around the lowest well-resourced scores. This trend suggests that despite population growth, infrastructure development has not kept pace in Sydney's western fringe.

3.2.1. Formula Rationale

To measure public transport accessibility at the SA2 level, we developed a composite score based on stop density and importance. Major stops were weighted more heavily, and the total was divided by each SA2's area to calculate stop density per square kilometre.

To ensure fair comparison across regions, we applied z-score normalisation to account for differences between urban and suburban areas. These z-scores were then transformed using a sigmoid function to scale values between 0 and 1, reducing the impact of outliers.

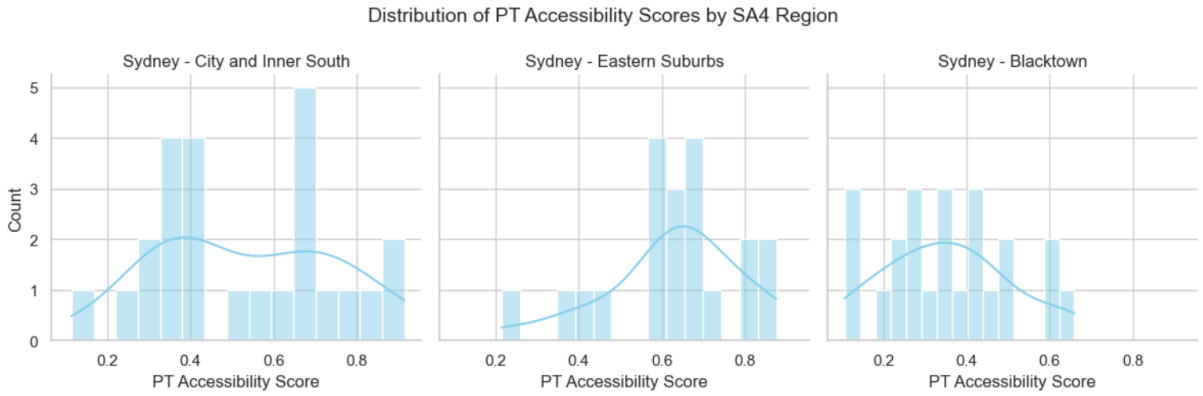
The final PT accessibility score reflects how well each SA2 is connected to the transport network, allowing consistent comparisons across varying geographies.

3.2.2. Impact of Different Components on Overall Score

The PT accessibility score is mainly driven by stop density per square kilometre, with higher weights for major transport hubs such as train stations. This means areas with more frequent and strategically placed services tend to score higher.
However, the number of stops isn't the only factor. Some regions with many stops still perform poorly if those stops are disconnected or inefficient. On the other hand, areas with fewer but better-connected stops can score well. The score also accounts for SA2 size, preventing larger areas from being unfairly penalised due to lower density.

This method ensures that the score reflects not just stop quantity, but also the quality of access and network integration.

3.2.3. Overall Distribution



Across the selected SA4 regions, PT accessibility scores reveal distinct spatial patterns. In the Eastern Suburbs, most SA2s fall within the moderate range (0.2–0.6), with a few high-scoring outliers, suggesting uneven access concentrated in inner areas. The City and Inner South displays consistently high scores (mostly above 0.6), supported by dense and multimodal transport infrastructure. In contrast, Blacktown shows a clear bimodal distribution—some SA2s score as high as 0.8, while others drop to around 0.3—highlighting significant infrastructure disparities within the region.This analysis highlights clear spatial disparities in public transport accessibility and resource availability across Sydney's SA4 zones. Inner-city areas benefit from dense, multimodal infrastructure, while outer suburbs face patchy access and under-resourcing. The findings underscore the need for targeted, SA2-level planning to address local infrastructure gaps and promote equitable urban development.

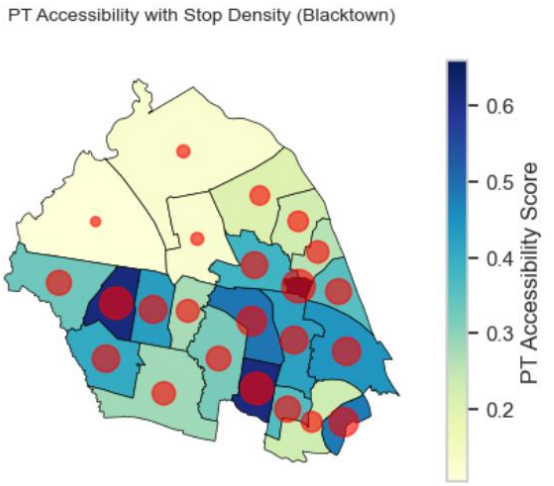
3.2.4. Comparison of Scores for Selected SA4 regions

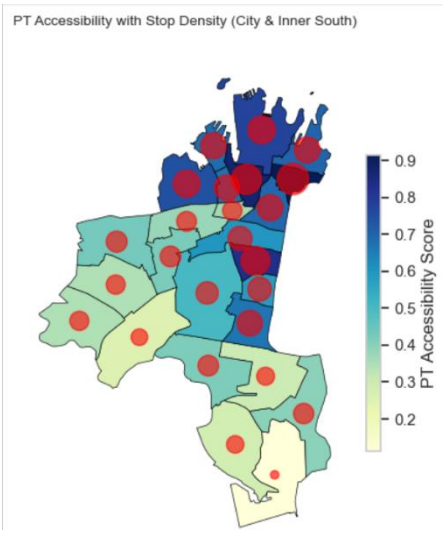
Maps of Public Transport Accessibility Scores across Blacktown, Eastern Suburbs, and City and Inner South reveal notable contrasts.

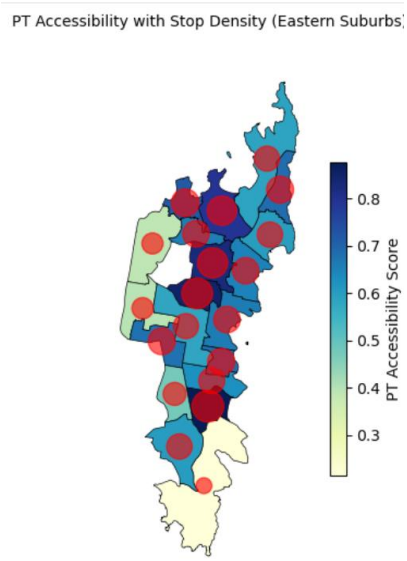
In the **Eastern Suburbs**, central and northern SA2s like Bondi Junction show high accessibility, while southern areas like La Perouse have low scores and sparse stop density.

In the **City and Inner South**, northern SA2s such as Surry Hills and Redfern have dense stop coverage and scores above 0.8, whereas southern zones like Mascot score below 0.4, showing a sharp contrast within a compact area.

Blacktown displays fragmented accessibility; a few central SA2s score around 0.6, but most outer areas fall below 0.5, indicating service gaps. Overall, these maps highlight strong accessibility in central zones and significant disparities in outer suburbs, reinforcing the need for more balanced transport planning.







3.2.5. Trends and Noteworthy Patterns

Several key trends emerge across the selected SA4 zones. Central areas like the City and Inner South consistently show high scores, thanks to their dense and multimodal transport networks. Most SA2s here score above 0.7, highlighting effective integration.

In contrast, the Eastern Suburbs and Blacktown show more variability. While a few areas perform well, many SA2s have low accessibility due to sparse or disconnected stop coverage. Blacktown, in particular, exhibits internal inequality—some SA2s near major stations score around 0.6, while others fall below 0.3.

Interestingly, a high number of stops doesn't always translate to better scores. In outer areas, red centroids may be present, but if they aren't well connected or frequent, accessibility remains low. This highlights that functional access—rather than just infrastructure presence—is key.

4.1. Correlation Analysis between Median Income and Well-Resourced Score

To explore the relationship between socioeconomic status and resourcing, a Pearson correlation was calculated between well_resourced_score and median_income across all SA2 regions. The result was a weak negative correlation of **-0.23**, suggesting that higher income does not necessarily translate to better public resourcing. (refer to appendix 4.1. for scatter plot and regression line.)

This outcome reflects the complexity of the well-resourced score, which incorporates multiple public service dimensions like health, education, and infrastructure, not all of which scale directly with income. In wealthier areas, reliance on private services may reduce public service presence, while lower-income areas might receive targeted public investment, boosting their scores. Regional planning and equity-driven policies may also direct more resources toward disadvantaged areas, further weakening the link between income and resourcing. Overall, income alone is not a strong predictor of how well resourced the regions are, reinforcing the need to consider a broader set of factors when assessing regional equity.

4.2. Correlation Analysis between

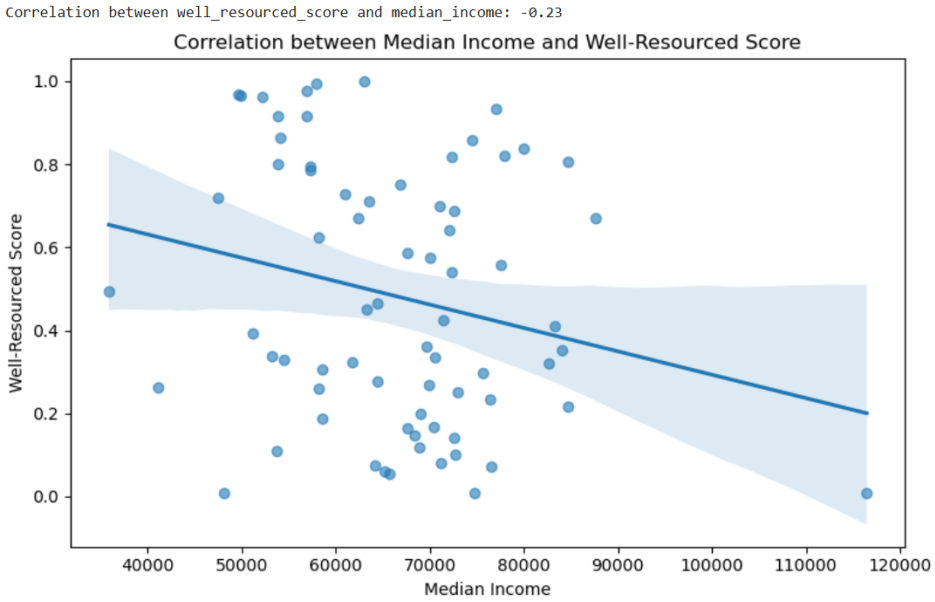
A correlation analysis between the public transport accessibility score and the well-resourced score found a coefficient of **-0.04**, indicating almost no linear relationship between the two.

This result challenges the assumption that better transport access leads to higher resourcing. It suggests that transport alone may not significantly impact overall resourcing unless paired with nearby services like schools, hospitals, or job centers. In some areas, good transport links may still lead to locations lacking essential public infrastructure or relying more on private alternatives.

The finding reinforces that resourcing is shaped by a combination of factors, and transport access is just one piece of a more complex picture.

APPENDIX: Defence of Approach

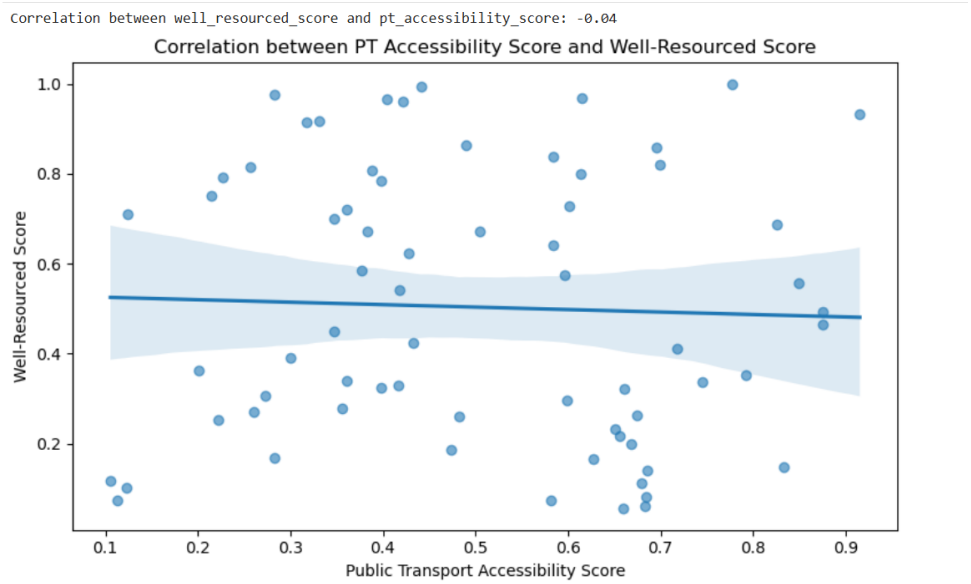
4.1. Correlation Analysis



Limitations of the correlation analysis:-

- Correlation does not imply causation – Even though there is a weak negative correlation of -0.23, this does not imply any causal effects between median income and well-resourced score. Other factors may have influenced both the variables.
- Index Complexity – The well-resourced score is a composite of several factors. This means that a weak correlation with median income may have masked stronger correlations between median income and other individual factors such as health services etc which have not ben separately analysed

4.2. Correlation Analysis



Limitation of the correlation analysis:-

- Correlation does not imply causation – Even though there is a very weak negative correlation of -0., this does not imply any causal effects between public transport accessibility score and well-resourced score. Other factors may have influenced both the variables.
- Limited Scope of PT Accessibility Metric - The PT accessibility score is based primarily on stop density and importance. It may not fully capture service frequency, reliability, or connectivity to critical destinations, which limits its influence on the overall well-resourced score.

REFERENCES

- Dodson, J., & Sipe, N. (2008). *Unsettling Suburbia: The New Landscape of Oil and Mortgage Vulnerability in Australian Cities*. Urban Research Program, Griffith University.
- SGS Economics & Planning. (2019). *Public Infrastructure for Growth: How Infrastructure Supports Australia's Regions*.
- NSW Department of Planning and Environment. (2022). *Greater Sydney Region Plan – A Metropolis of Three Cities*.