DATA2001 Lab 05 Group 02

SID540028959 SID540234273 SID530333591

1. Dataset Description

1.1 Data Sources

Data	File Name	Type	Key Columns / Info	Spatial?
businesses	Businesses.csv	DataFrame	'industry_name', 'sa2_code', 'total_businesses'	No
income	Income.csv	DataFrame	'sa2_code21', 'median_age'	No
population	Population.csv	DataFrame	'Sa2_code', age groups, 'total_people'	No
stops	stops.txt	DataFrame →GeoDF	'stop_id', 'stop_lon', 'stop_lat'	Yes (converted)
school_future	catchments_future.shp	GeoDataFrame	'USE_ID','geometry'	Yes
school_primary	catchments_primary.shp	GeoDataFrame	'USE_ID','geometry'	Yes
school_secondary	catchments_secondary.shp	GeoDataFrame	'USE_ID','geometry'	Yes
sa2	SA2_2021_AUST_GDA2020.shp	GeoDataFrame	'SA2_CODE21', 'SA2_NAME21', 'SA4_NAME21', 'GCC_NAME21', 'geometry'	Yes

NSW Points of Interest API:
https://maps.six.nsw.gov.au/arcgis/rest/services/public/NSW_POI/MapServer

1.2 Pre-processing Steps

-SA2 filtering by Greater Sydney

```
sa2=sa2[sa2["GCC_NAME21"]=="Greater Sydney"]
print(sa2["SA4_NAME21"].unique())
```

```
['Central Coast' 'Sydney - Baulkham Hills and Hawkesbury'
 'Sydney - Blacktown' 'Sydney - City and Inner South'
 'Sydney - Eastern Suburbs' 'Sydney - Inner South West'
 'Sydney - Inner West' 'Sydney - North Sydney and Hornsby'
 'Sydney - Northern Beaches' 'Sydney - Outer South West'
 'Sydney - Outer West and Blue Mountains' 'Sydney - Parramatta'
 'Sydney - Ryde' 'Sydney - South West' 'Sydney - Sutherland']
```

-Spatial transformation

DATA	SRID	GEOM.TYPE
schools	3857	Polygon→MULTIPOLYGON
sa2	4326 / 3857	Polygon→MULTIPOLYGON
stops	4326	POINT

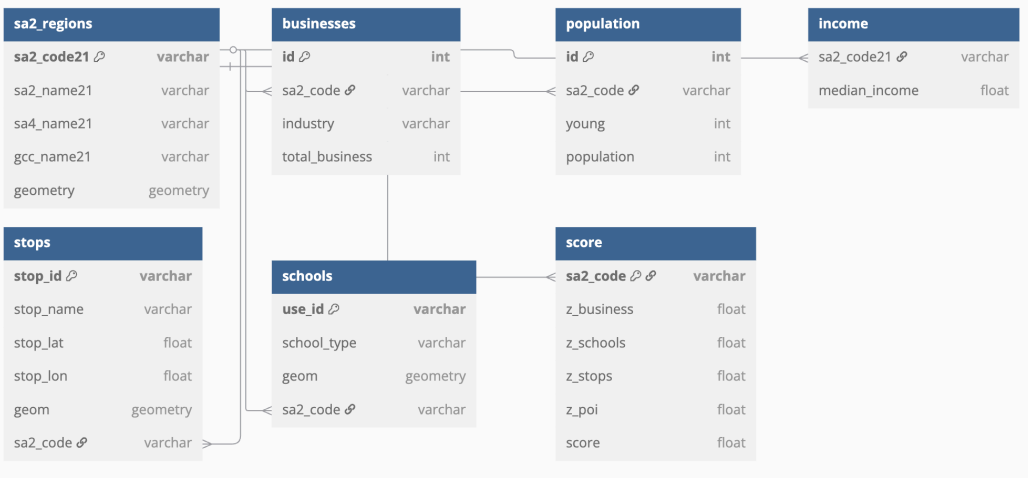
So we chose the SA4 regions of 'Sydney - Inner West', 'Sydney - Blacktown', 'Sydney - Parramatta'.

2. Database Description

2.1 Schema Design

schools = school_future **UNION** school_primary **UNION** school_secondary

population['young'] = population['0-4_people'] + population['5-9_people'] + population['10-14_people'] + population['15-19_people']



2.2 Data Integration

valid_sa2	1. Filter out SA2 areas within selected SA4 2. The population is ≥100
business_per_1000	1. Joining **population** and **businesses** on **sa2_code.** 2. Group by **industry (industry_name)** 3. Calculate the number of **total_business** per 1000 people
stop_count_per_sa2	1. Use **ST_Within** to spatially connect the bus stop with the SA2 boundary (**SRID:4326**) 2. Group by SA2 and count the number of stops
school_catchments_per_1000_youth	1. Using **ST_Intersects** to connect with SA2 2. Using **ST_Area** to calculate the catchment areas per 1000 young
poi_count_per_sa2	1. Aggregate the point of interest data returned by the NSW API. 2. Group and count by the SA2 region

2.3 Indexing

```
CREATE INDEX IF NOT EXISTS idx_sa2_geom
ON sa_2
USING GIST (geom);

CREATE INDEX IF NOT EXISTS idx_sa2_code
ON sa_2 (sa2_code21);
```

idx_sa2_geom: Spatial index on the geom column using the GiST structure. This improves the performance of spatial operations such as ST_Within and ST_Intersects
idx_sa2_code: Index on the sa2_code21 column for fast attribute joins with other tables that use sa2_code as a key.

3. Score Analysis

3.1 Scoring Formula

Score = *S*(*z*-business + *z*-stops + *z*-schools + *z*-POI)

S is the sigmoid function, *z* is the normalised z-score

z_business	Businesses per 1000 people, in selected industries
z_stops	Number of public transport stops
z_schools	Catchments areas per 1000 'young people' (0-19 ages)
z_poi	Number of places of interest, in selected groups

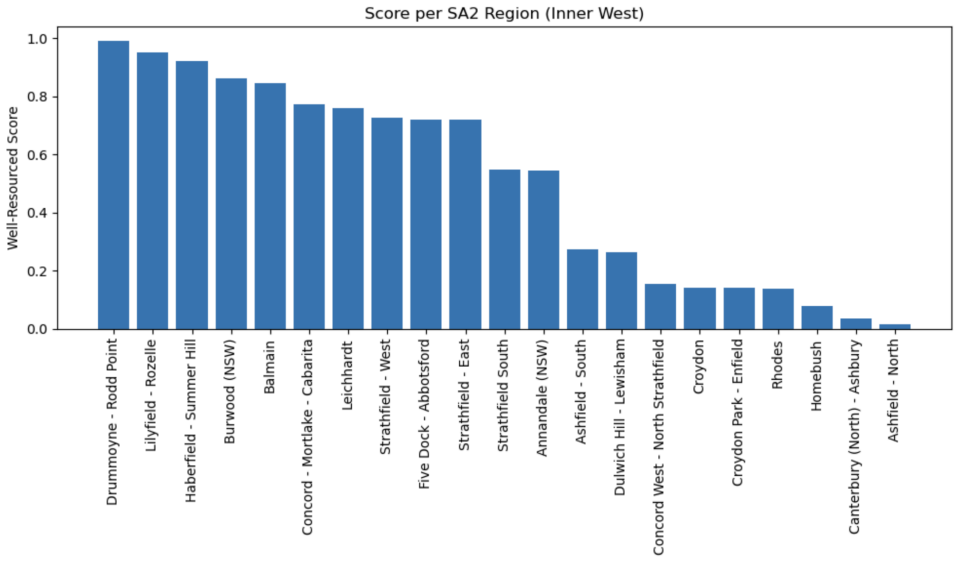
z_business (industry_name)		**z_poi (poi_group)**	
These sectors reflect essential **local services**, commonly found in **community-focused areas**, and represent both **economic activity** and access to everyday resources.	**Education and Training**	These categories capture essential **non-commercial services** that contribute to a region's overall **livability and infrastructure**.	**1: "Community"**
	Health Care and Social Assistance		**2: "Education"**
	Professional, Scientific and Technical Services		**3: "Recreation"**
	Accommodation and Food Services		**4: "Transportation"**

First, we conduct Z-score standardization on the four resource indicators (business, stops, schools, POI) to obtain the relative performance values of each indicator. Then, we sum the four standard scores and map them to the [0, 1] interval through the Sigmoid function to form the final score. This method can reduce the influence of different dimensions on the scoring, making different indicators comparable in the scoring.

$$\sigma(x) = \frac{1}{1+e^{-x}} = \frac{e^x}{1+e^x} = 1 - \sigma(-x). \qquad z = \frac{x-\mu}{\sigma}$$

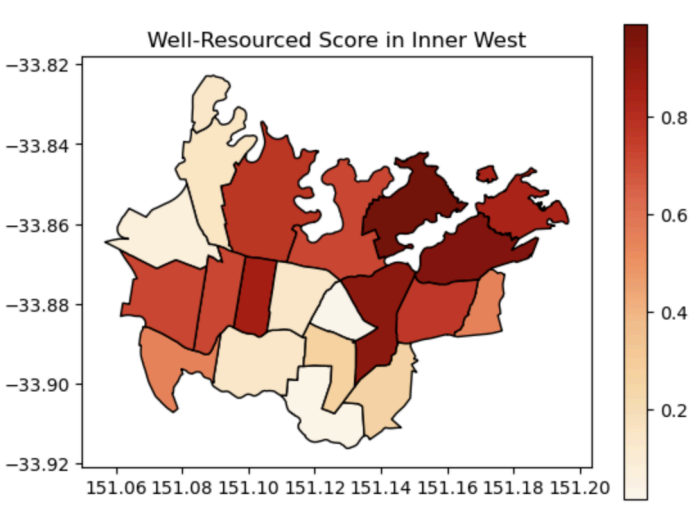
3.2 Score Distribution

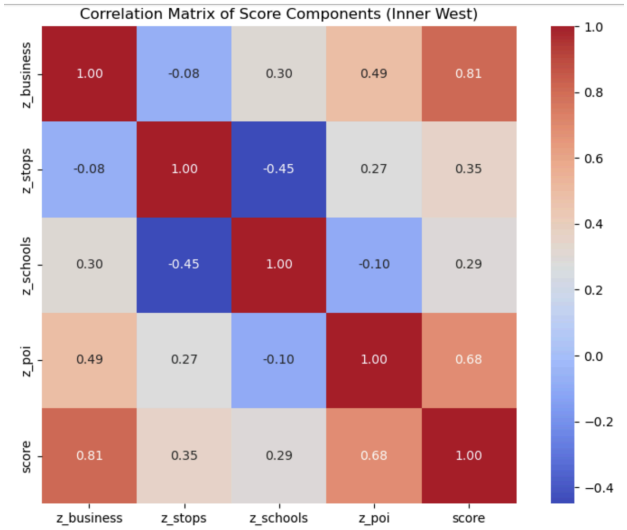
-Inner West (SID540028959)



The bar chart highlights the significant differences in resource adequacy scores between SA2 and SA4 in the Inner West. Areas like Drummoyne - Rodd Point and Lilyfield - Rozelle are resource-rich, with strong access to businesses, schools, transport and attractions. In contrast, areas like Ashfield - North and Ashbury have significantly lower scores, reflecting limited accessibility.

This shows that resource accessibility is not evenly distributed, even within a relatively small and central urban area of Sydney.

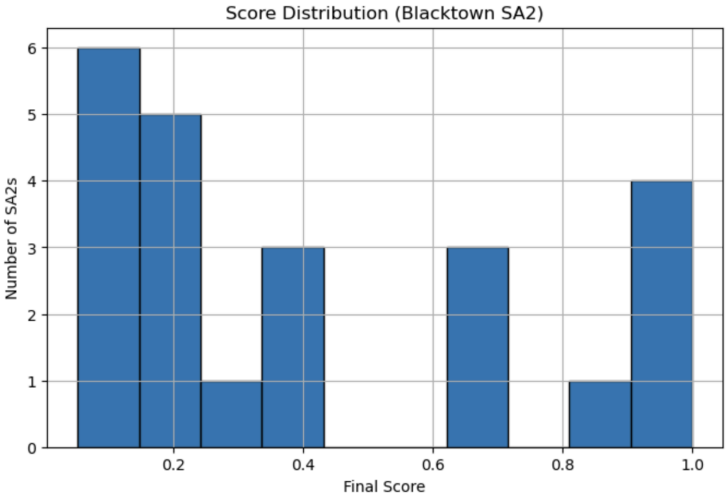


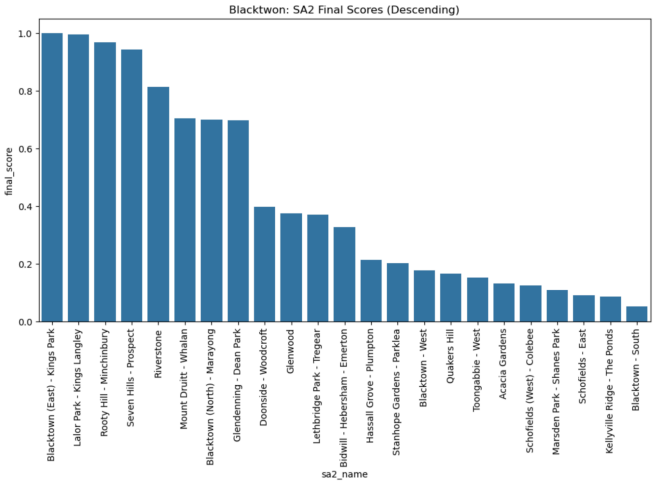


Choropleth Map:
The map reveals spatial inequality across the Inner West. Northern SA2 regions like Drummoyne - Rodd Point and Burwood (NSW) are resource-rich, while southern areas such as Ashfield - North and Ashbury have fewer combined resources, as reflected by lighter shading.

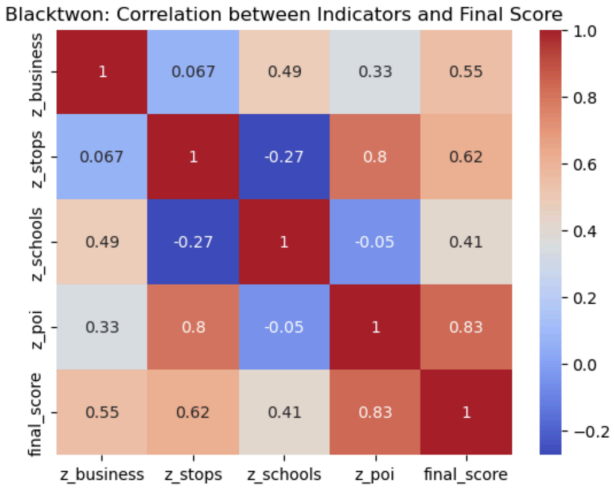
The heatmap shows that z_business (r = 0.81) and z_poi (r = 0.68) are the strongest contributors to the final score. In comparison, z_stops and z_schools have weaker relationships, suggesting that access to transport and school catchments does not always align with other indicators.

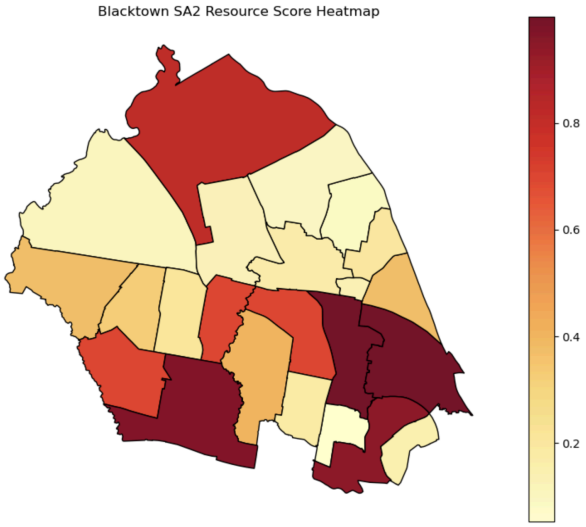
-Blacktown (SID540234273)





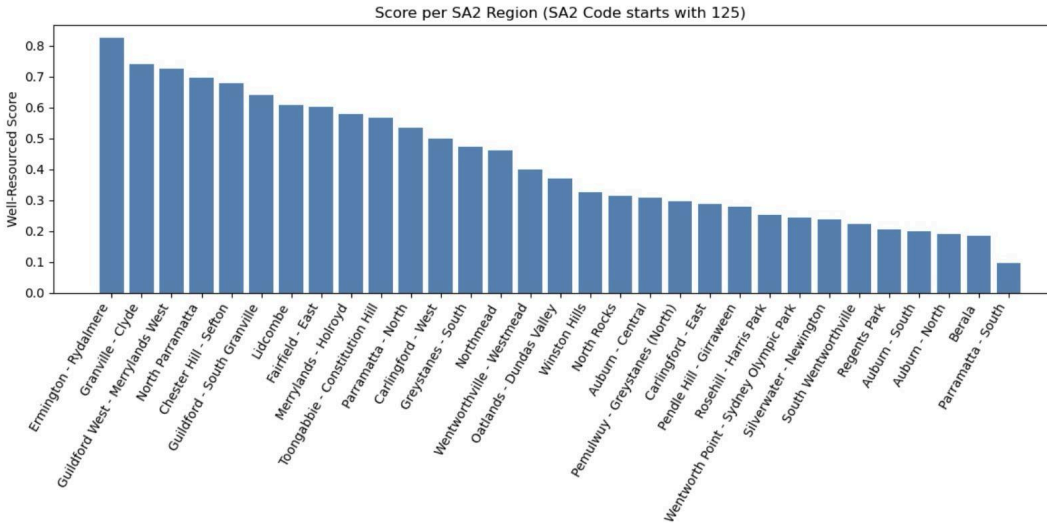
The Score Histogram shows that most of the SA2 scores are concentrated in the range of 0.1 to 0.3. Only a few high-resource areas have scores close to 1, and the distribution leans to the left, showing a positive bias Distribution of Final Resource Scores.



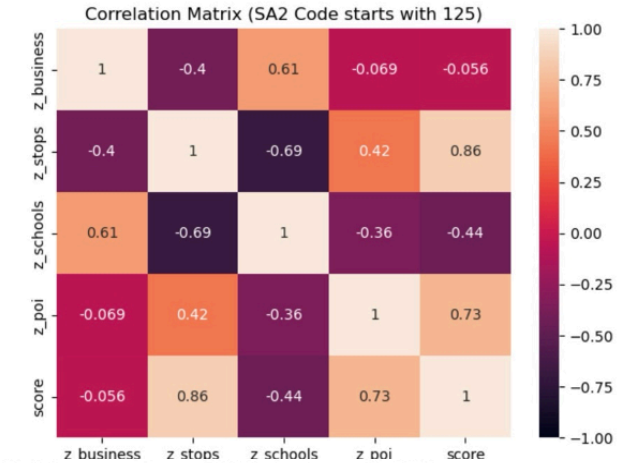


We visualised the Pearson correlation coefficients
between each individual indicator (z-score) and the final score. As shown in the heatmap, the number of POIs has the strongest positive correlation with the final score (0.83), followed by public transport stops (0.62). This indicates that these two indicators played the most significant role in the final resource scoring within the selected SA4 region.
This choropleth map visualises the final resource scores for each SA2 within the Blacktown SA4 region. The deeper red areas represent higher scores, indicating more well-resourced SA2s, while lighter regions suggest fewer available resources relative to the metrics used.

-Parramatta (SID 530333591)



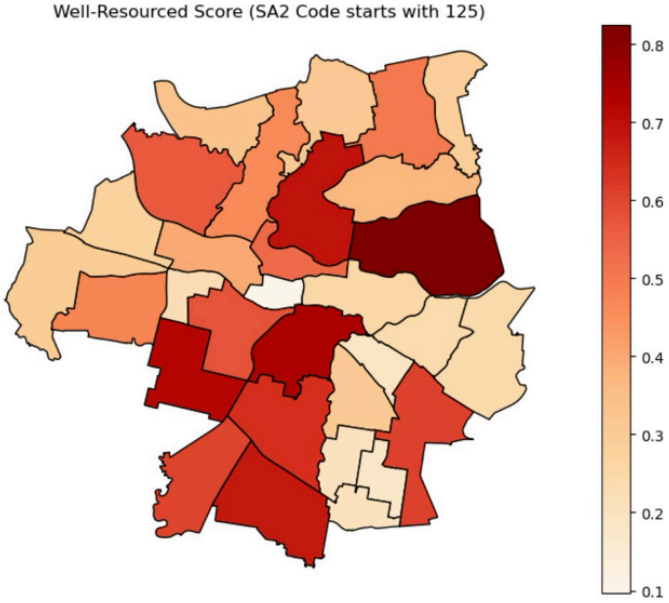
Bar Chart Clearly illustrates a “long-tail” distribution: a small number of SA2s have outstanding resource access, The score between 0.8-around 0.15, while many others trail behind. High scores tend to coincide with a balance of strong transport, business, education, and POI (points of interest) metrics.



The correlation between school density (z_schools) and the overall score is moderate (correlation = -0.44), and its sign is negative. This suggests that, in this specific region, having more schools is not always associated with higher resource access. It may reflect that some areas with high school density are not as well-resourced in other aspects.

The business metric (z_business) has the weakest correlation with the final score (correlation = -0.056), indicating it plays a smaller role in driving the overall resource score in this dataset.

The heatmap shows that access to POIs and public transport are the decisive factors for overall resource access in the Parramatta (125) region, while schools and business presence play a less significant or even slightly negative role in determining the score.

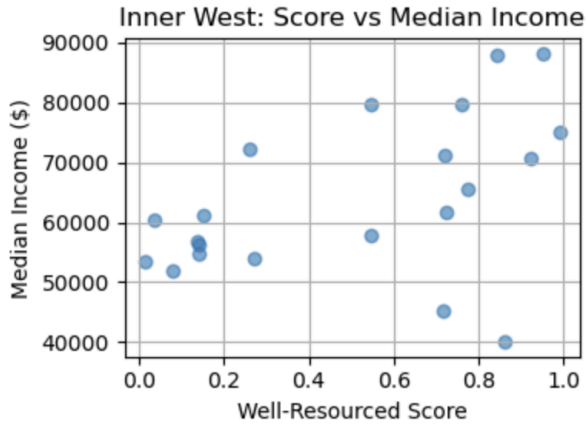


This pattern is consistent with the urban structure of Western Sydney: areas closer to the CBD and major transport nodes (e.g., Parramatta, Granville, Ermington) benefit from infrastructure concentration, while further suburbs may be underserved.

The map provides a clear, intuitive visual summary of spatial inequality in resource access within the Parramatta region. It highlights the need for targeted planning and investment to improve conditions in lower-scoring suburbs.

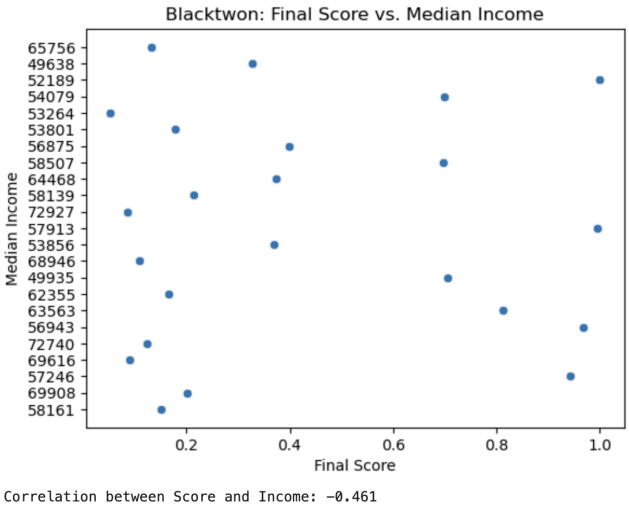
4. Correlation Analysis

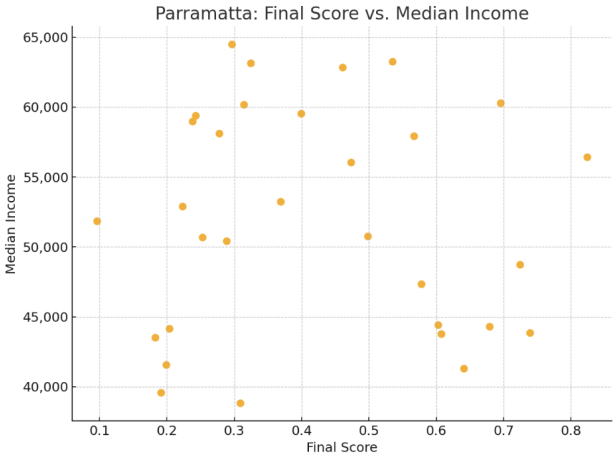
4.1 Method



A Pearson correlation test between the well-resourced score and median income across Inner West SA2s yielded a coefficient of r = 0.463, suggesting a moderate positive correlation. This indicates that SA2 regions with greater access to public resources tend to have higher income levels. However, the relationship is not especially strong, as several low-scoring areas report moderate-to-high incomes, and some high-scoring regions do not rank among the most affluent.

This scatter plot explores the relationship between the score in Blacktown and the median income. It shows a moderate negative correlation (r = -0.461), indicating that some regions with better resources might be associated with those with lower incomes. This discovery may reflect the existence of public infrastructure in low-income communities.





This scatter plot explores the relationship between the score in Parramatta and the median income. It shows little correlation (r = –0.027), This indicates that there is no significant linear relationship between Final Score and Median Income. This indicates that wealth distribution in the region is relatively dispersed.

4.2 Result

The result highlights that resource access and wealth do not always align perfectly, likely due to mixed land use, diverse population profiles, or historical infrastructure patterns within the region.

4.3 Limitations & Final Thoughts

1. Our score does not account for quality of services, only quantity (e.g., number of stops or businesses).
2. All features were equally weighted — changing weights could affect correlation.