Project 2: Exploring the Profiles of Melbourne's Suburbs

CP217: Machine Learning for Cyber-Physical Systems August-Dec Semester 2024

Team: Saferide Squad

Vanshika Jindal-24721, Vandana Mourya-22969, Guha H-24639

1 Part A: Hypothesis-Driven Research

1.1 : Similarity Measures

Similarity Measure 1: Geographical measure aims to evaluate how similar 2 suburbs are based of geographical features - location, map reference and grid reference. These features are selected because they represent suburb location with respect to Melbourne. Method used to combine features: Location parsing (extracts distance and direction, calculates latitude and longitude using Haversine formula), Euclidean distance (used for map and grid reference) and Haversine formula (great-circle distance between all suburb coordinates).By combining Haversine and Euclidean distances, we compute a final geographical similarity score by averaging both measures. This is baseline similarity measure used for further analysis.

Similarity Measure 2: Demographic Similarity of Suburbs (Population Percentage of 64+ Years), this measure evaluates suburban similarity using features - population percentage of 64+ years in 2007 and 2012. Features are chosen to reflect changes in the aging population, which impact healthcare, housing, and services. Method for Combining Features: Feature Extraction (Extract the percentage of the population aged 64+ in 2007 and 2012), Distance Calculation (Compute Manhattan distance between both features for each suburb pair to create a distance matrix). Using Manhattan distance helps identify neighborhoods with similar aging trends by focusing on absolute differences in percentages. This demographic measure integrates data with spatial representation.

Similarity Measure 3: This measure compares suburbs based on cultural background, linguistic diversity, and migration patterns using these features: Top Country of Birth, Top Languages Spoken, Aboriginal or Torres Strait Islander status, Born Overseas, and Speaks LOTE (Language Other Than English) at Home. Metric used: Jaccard Similarity (Used for categorical data to measure overlap), Euclidean Distance (Used for continuous features to measure numerical differences).Both measures are normalized to [0,1], then combined for a comprehensive similarity score. This approach captures cultural and demographic diversity in suburbs, incorporating both categorical and numerical features.

1.2 : Visualize Similarities

Similarity Measure 1: In the MDS plot (Fig. 1), nearby Melbourne suburbs like North Melbourne, South Melbourne, and Northcote cluster together, reflecting their real-life proximity. Distant suburbs, like Somerville and Sorrento, appear farther apart, aligning with their physical separation. Overall, the plot meets expectations, though slight discrepancies exist—for example, Glenroy appears equidistant from Fawkner and Braybrook, despite being closer to Fawkner in reality.

Similarity Measure 2 : In the MDS plot (Fig. 2), suburbs with similar elderly populations, like Fawkner, Glenroy, and Noble Park, cluster together, while areas like Melbourne Airport and Sorrento appear distant, indicating unique age demographics. Some geographically close inner suburbs, such as St Kilda and Windsor, show separation in aging trends, highlighting demographic diversity even among neighboring areas.

Similarity Measure 3 : Suburbs with high cultural diversity, such as North Melbourne, St Kilda, and Footscray, cluster together in the MDS plot (Fig. 3), indicating similar cultural and linguistic demographics. Quantitative analysis of cultural diversity can be observed in Fig. 13.

Conclusion: South Melbourne, South Yarra, St Kilda, St Kilda East remain consistently close across all plots hence similar under each measures.

1.3 Hypothesis-Testing :

To get a better understanding of actual location of Suburbs we have simulated actual map of Melbourne (Fig. 7).

Suburbs near each other, like South Melbourne, South Yarra, St Kilda, St Kilda East, cluster (number of clusters chosen on basis of Silhouette score) din the MDS plot (Fig 4, Fig. 5, Fig. 6), supporting the idea that close suburbs are often similar due to shared traits. Some close suburbs, like Toorak and St Kilda East, do not cluster together, likely due to map/grid reference influence, showing that proximity alone doesn't always determine similarity.Geographic proximity was estimated by averaging Haversine (real-world distance) and Euclidean (grid-based spatial relationships) distances, providing a comprehensive similarity measure.

2 Exploratory Data Analysis

2.1 Identify Patterns

The analysis reveals a stark divide between regions near Melbourne city and those located farther away. Suburbs like Somerville, Tyabb, Sorrento, and St Andrews Beach are geographically distant from Melbourne and show underdevelopment in several key areas (Fig. 11) These regions tend to have lower percentages of degree holders, a higher proportion of residents with lower incomes (Fig. 9), and a greater number of people who did not complete year 12 compared to more urbanized areas. In terms of population dynamics, Sorrento stands out with a notably higher elderly population compared to younger residents, which highlights the aging demographics in this suburb (Fig 2). Additionally, the data points to a significant variation in population growth across the suburbs, with some experiencing substantial increases (e.g., Braybrook) while others, such as some of the aforementioned suburbs, have seen little to no growth or even a decline (Fig.14). Healthcare access (Fig. 12) also varies widely, with areas like Parkville and St Kilda West facing significant challenges due to elderly and disability vulnerability, compounded by transportation issues. Vulnerability scores indicate that regions like Moorabbin and St Kilda East have high elderly vulnerability, while areas such as Pascoe Vale South and Sorrento have higher disability vulnerabilities (Fig. 10)

2.2 Explain Observations

The underdevelopment in suburbs like Somerville, Tyabb, Sorrento, and St Andrews Beach can largely be attributed to their distance from Melbourne's city center, which limits their access to essential services, jobs, and infrastructure (Fig. 7). These suburbs exhibit lower levels of education attainment, with fewer degree holders and a higher percentage of residents who have not completed year 12, suggesting limited access to higher education opportunities. Additionally, these regions have lower income levels, indicating fewer economic opportunities. Sorrento's higher elderly population can be linked to the region's appeal as a retirement destination, which often attracts older residents. The lack of development in these areas also impacts healthcare access, as evidenced by the vulnerability scores, which show these regions struggling with both elderly and disability-related challenges. The overall lack of transport options further exacerbates the difficulties faced by residents in accessing vital services.

2.3 Reflect on Insights and Benefits

Eyeballing the raw data would not reveal these complex relationships between socio-economic, healthcare, and demographic factors. Without the use of clustering techniques, vulnerability scoring, and detailed visualizations, the interconnectedness between population growth, socio-economic disadvantage, healthcare access, and transport limitations would be missed. These patterns only become apparent through structured analysis, which helps identify which regions require attention and resources the most. For example, the need for targeted healthcare interventions in vulnerable suburbs would not be obvious without such analytical methods.

The insights gained from this analysis and the correlation between various socioeconomic features (Fig. 8) can drive meaningful improvements in public policy and infrastructure. For example, targeting healthcare infrastructure in areas with high elderly or disability vulnerability, such as Parkville and Moorabbin, could improve service access. Similarly, enhancing transport options in remote areas like Melbourne Airport and St Andrews Beach would ensure better access to services for residents. Addressing these disparities would contribute to more equitable access to essential services, reduce social inequality, and improve the overall quality of life for people in underserved areas, ultimately benefiting the community at large.

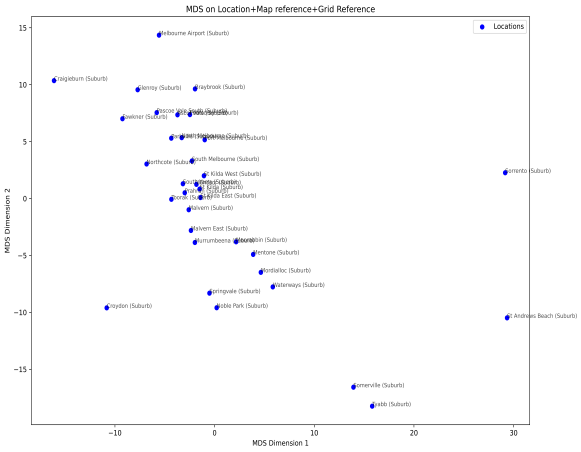

Figure 1: Similarity Measure 1

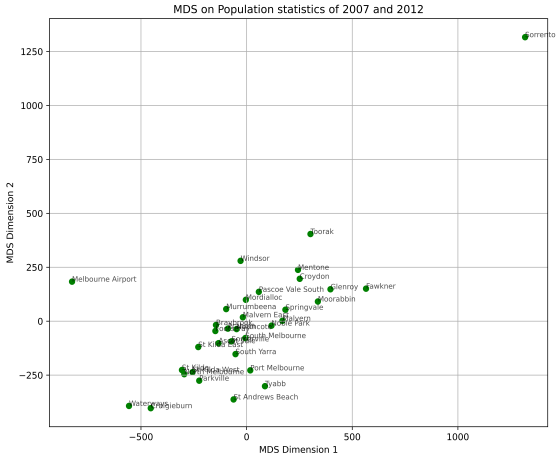

Figure 2: Similarity Measure 2

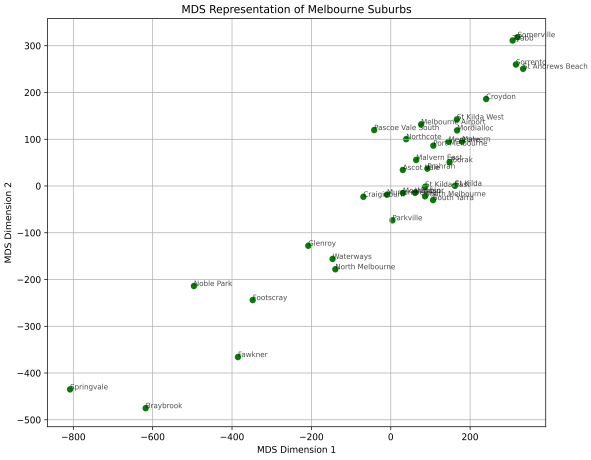

Figure 3: Similarity Measure 3

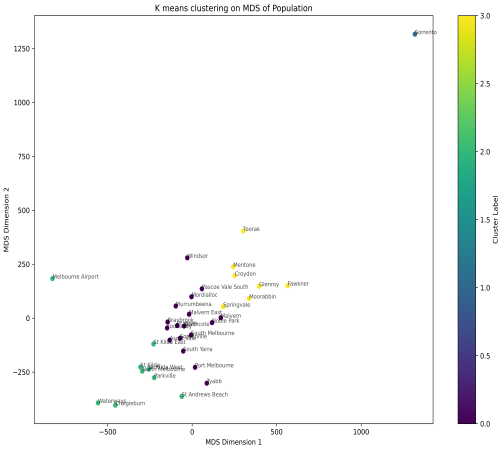

Figure 4: K means clustering of Similarity Measure 1

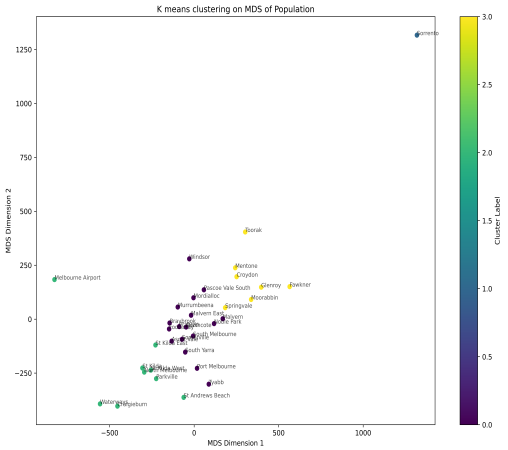

Figure 5: K means clustering of Similarity Measure 2

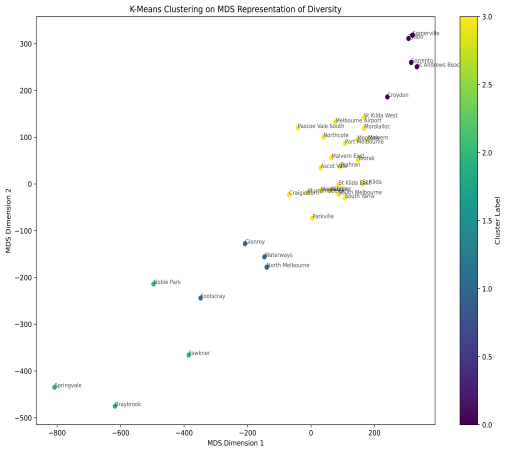

Figure 6: K means clustering of Similarity Measure 3

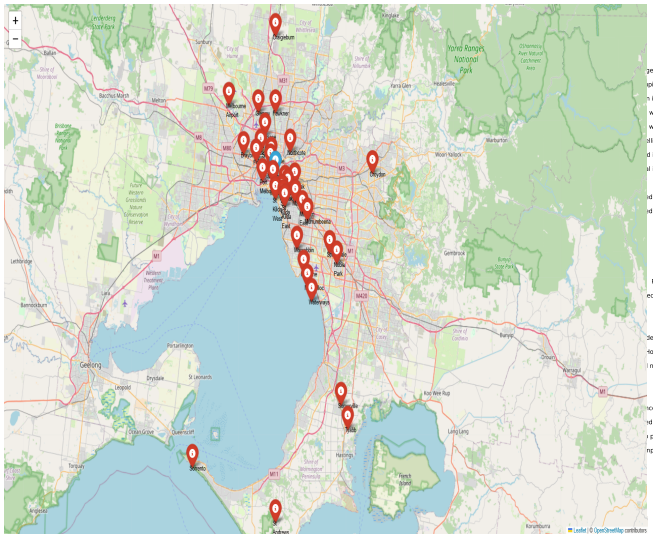

Figure 7: Actual map of melbourne

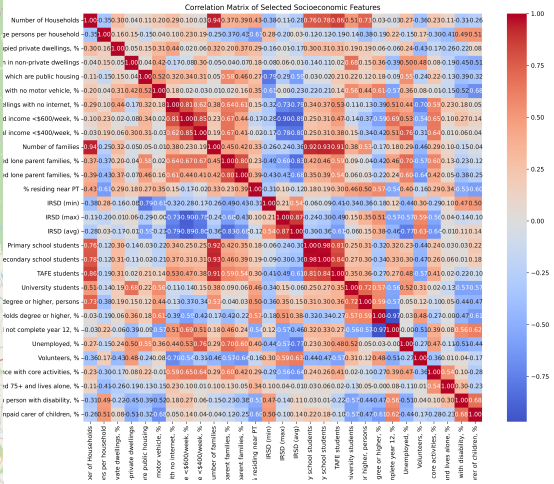

Figure 8: Correlation matric of socioeconomic features

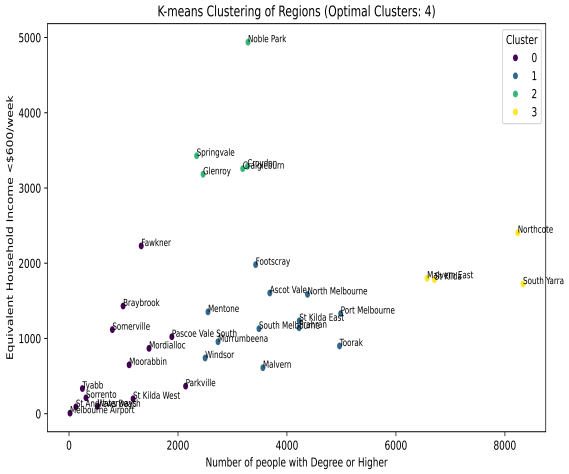

Figure 9: K means clustering on income vs education plot

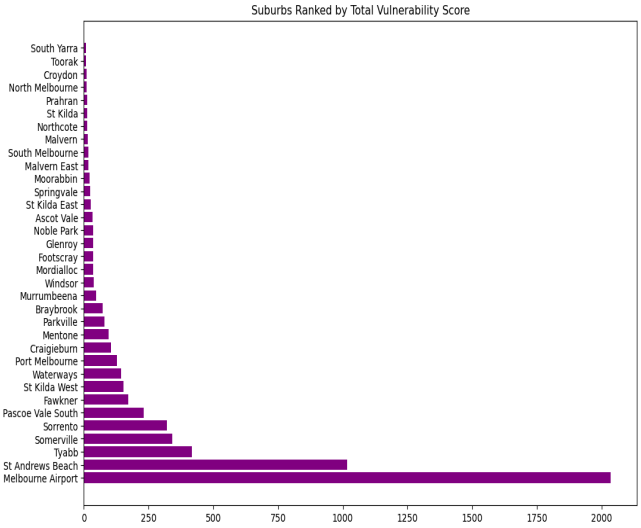

Figure 10: Suburbs ranked based on vulnerability score

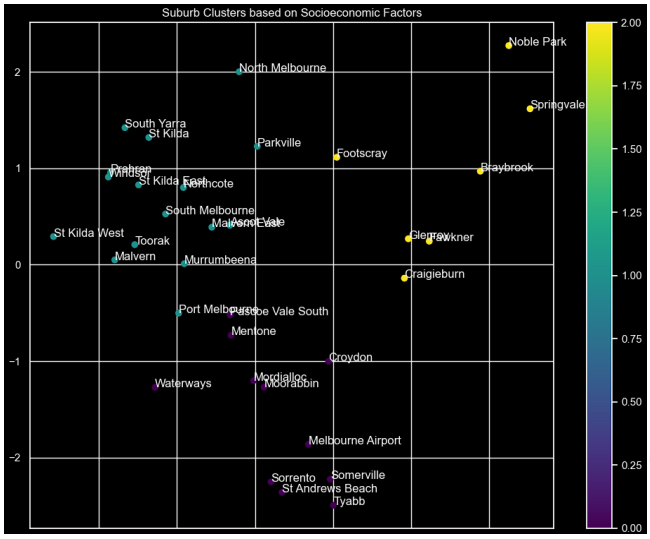

Figure 11: K means clustering based on Socio-demographic features

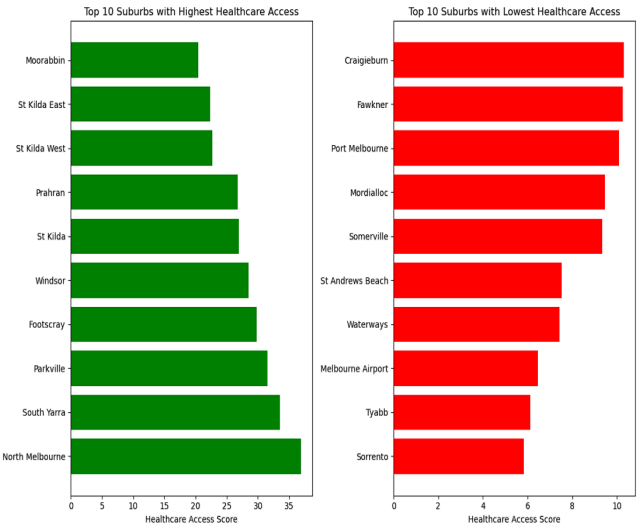

Figure 12: Ranking of Suburbs based on healthcare access

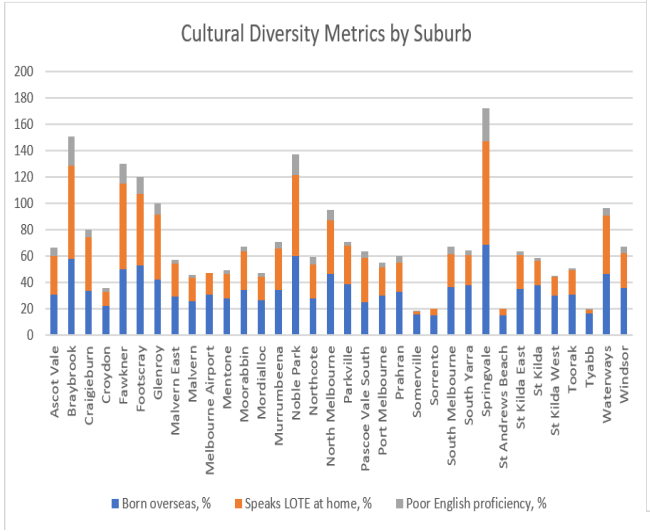

Figure 13: Cultural Diversity Across Regions

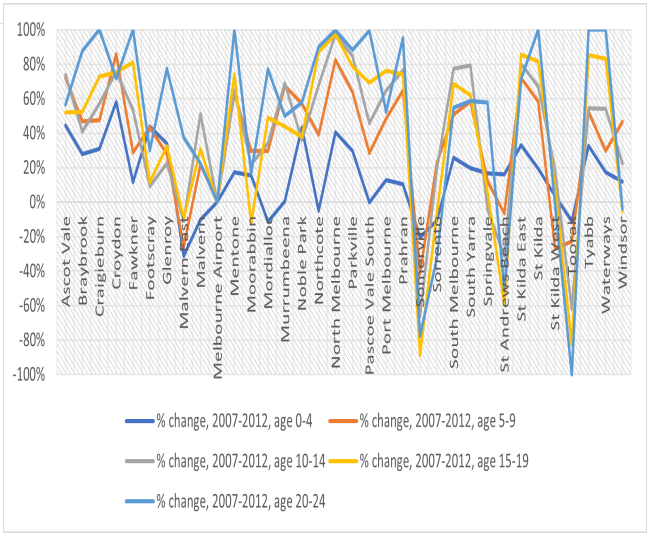

Figure 14: Percentage change in Population(2007-2012)