Forecasting Aged Care Facility Demand in Australia

Hazim Shaikh
School of Computer and Mathematical Sciences
The University of Adelaide
Adelaide, Australia
a1894307@adelaide.edu.au

***Abstract*—The rapidly aging population of Australia is posing serious challenges for the elderly care industry. As of 2023, 4.4 million Australians—around 17% of the total population—were aged 65 or older Australian Bureau of Statistics (2023*b*), and this figure is projected to rise to nearly 25% by 2050 Productivity Commission (2011*a*). This study harnesses big data—combining national aged care service datasets and demographic projections—to forecast demand for aged care facility capacity across Australia's regions. We apply a multi-stage approach incorporating data integration, clustering of regions by service patterns, predictive regression modelling, and scenario-based forecasting through 2032. K-means clustering reveals distinct regional profiles of aged care provision (e.g., urban versus remote patterns). Regression models (linear and machine-learning) achieve near-perfect fit (e.g., Linear R^2 = 1.000, Ridge $R^2 \approx$ 0.999997) with minimal error, indicating a strong relationship between elderly population and service capacity. Using official population projections, we forecast significant growth in needed residential care places nationally—potentially tens of thousands of additional beds by 2032—with the greatest shortfalls in fast-ageing coastal and rural areas. This report offers a novel, data-driven lens on aged care planning, with implications for resource allocation and policy to ensure equitable care for Australia's growing elderly population.**

***Index Terms*—component, formatting, style, styling, insert**

I. Introduction

Australia's population is aging at an extraordinary rate, and the country is on the verge of an aged care disaster.

Background and Objective: Due to longer lifespans and fewer births, the number of elderly Australians is increasing. In 2023, more than 4.4 million Australians, or roughly 17% of the population, were 65 years of age or older (Australian Bureau of Statistics, 2023*b*). By mid-century, nearly one in four Australians will be 65 or older (Australian Bureau of Statistics, 2023*b*; Productivity Commission, 2011*b*), and demand for aged care is projected to triple compared to current levels (Productivity Commission, 2011*b*). This "grey tsunami" is already straining aged care services – currently around 1.3 million Australians access some form of aged care, and about 5% of people over 65 reside in nursing homes (Productivity Commission, 2011*b*). The sector has been declared in "national crisis" due to workforce shortages and service shortfalls (Productivity Commission, 2011*b*).

A crucial public policy concern is making sure that sufficient residential aged care facilities (such as nursing home beds and assisted living facilities) are accessible when and where needed. However, varying care preferences (e.g., many seniors choose to age at home) and unequal regional ageing patterns make designing such capacity challenging. The need for data-driven forecasting to predict future demand for elderly care facilities throughout Australia's various regions and support proactive capacity planning is what spurred this research.

Research Question: Which regions of Australia are likely to face the greatest increases in demand for aged care facility capacity over the next decade, and can we predict these trends using integrated big data on demographics and current service provision? In particular, we ask how projected growth in the elderly population (by small area) will translate into required aged care facility places, and whether current infrastructure is sufficient or if gaps will emerge. The analysis focuses on forecasting the demand for residential aged care facilities (including nursing homes and related services) by 2032, across statistical regions, using a combination of clustering and predictive modeling.

Contributions: This research makes several contributions:

- **Big Data Integration:** We compile and merge large-scale datasets from the Australian Institute of Health and Welfare (AIHW) (Australian Bureau of Statistics, 2023*a*) and Australian Bureau of Statistics (ABS) (Australian Bureau of Statistics, 2023*b*) – including a national aged care services dataset and population projections – creating a rich data foundation spanning all Australian regions.
- **Clustering of Regions:** We perform unsupervised clustering to identify patterns in aged care service provision by region. This reveals distinct regional profiles (e.g., metropolitan vs. remote areas) in terms of aged care capacity and service mix, providing insight into geographic disparities.
- **Predictive Modeling:** We develop and evaluate multiple regression models (including linear regression, Ridge regularisation, and ensemble methods) to quantify the relationship between demographic features (primarily the elderly population size) and current aged care capacity. Our models achieve extremely high explanatory power (R^2 near 1.0), highlighting a robust predictive link that we leverage for forecasting.
- **Demand Forecasting:** Using the best-performing model and ABS demographic projections, we forecast future aged care facility demand (in number of places required) for each region up to 2032. We identify potential short-

This work was completed as part of COMP SCI 7319OL – Big Data Analytics.

falls in capacity, highlighting which regions will likely need the most additional resources.
- **Policy Insight:** We provide an evidence-based discussion of the implications of these findings, particularly how population ageing will impact metropolitan versus rural and remote areas differently. This includes interpreting service gaps in the context of Australia's remoteness classification and discussing strategic responses (e.g., targeting infrastructure investment or bolstering in-home care where appropriate).

Through these contributions, our work demonstrates the power of big data analytics to inform one of Australia's most pressing social challenges – planning for adequate aged care in an ageing society. The approach and findings can help policymakers, aged care providers, and communities proactively prepare for the impending surge in demand.

II. Literature Review

Forecasting aged care demand has been a focus of both academic studies and government inquiries, with varied methodologies. We review five relevant works (summarised in Table I) and contrast them with our approach. Prior studies have examined ageing demographics, service utilisation, and capacity needs from different angles, including macro-level projections, regional analyses, and policy framework evaluations.

Nationwide Projections: The Productivity Commission's landmark *Caring for Older Australians* report Commission (2011) projected that by 2050 over 3.5 million Australians would be using aged care services annually—more than triple the number in 2010—and that the sector would require approximately one million paid workers to meet this demand. This high-level government forecast highlighted an impending surge in demand and underpinned major policy reforms, but it did not granularly map where capacity needs would be most acute. Our work builds on such macro forecasts by introducing a spatially detailed prediction across local regions.

State-Level Trend Analysis: Austin et al. Austin et al. (2021) examined aged care needs in New South Wales, projecting that demand would reach 135% of the 2019 capacity by 2029—implying a shortfall of over 25,000 residential care places. Their analysis, based on extrapolating utilisation trends against population growth, sounded an early warning for NSW. We extend this idea nationally, using a data-driven model to identify similar capacity gaps emerging across other states and regions.

Rural and Remote Focus: Blackberry and Morris Blackberry and Morris (2023) employed a time-series analytical approach to assess rural aged care needs. Using ABS and AIHW data, they quantified an existing shortfall of over 2,000 residential care places in rural and remote areas as of 2021, and projected an additional 3,390 places would be required by 2032. Their findings confirm significant geographic disparities—an insight our clustering analysis also uncovers—and reinforce the importance of targeted planning for non-urban areas.

Capacity Planning Ratios: Cooper-Stanbury Cooper-Stanbury (2025) investigated the distribution of aged care supply relative to planning benchmarks. That study revealed a shortfall of nearly 25,000 beds nationwide in 2023, projected to rise to 55,000 by 2033 without intervention. Australia's historical ratio of 78 places per 1,000 people aged 70+ was also revised downward as part of aged care reform. Our study uses a machine learning approach rather than fixed planning ratios, learning demand from actual data and forecasting regional-level deficits with greater precision.

Aged Care Service Models: Lewis et al. Lewis et al. (2025) critiqued existing aged care delivery models, especially the limitations of in-home care. Despite policy interest in "ageing in place," their analysis showed workforce and funding challenges make this difficult to scale. This supports our assumption that facility-based care will remain crucial and that forecasting facility capacity is still highly relevant.

Discussion of Literature: Overall, prior research agrees that Australia's ageing population will drive a dramatic increase in aged care demand in coming decades. Government reports and academic studies have underscored the scale of the challenge—from national projections of tripled service usage Commission (2011) to state-specific warnings of shortfalls Austin et al. (2021) and rural service inequities Blackberry and Morris (2023). However, there is a gap in granular, region-specific forecasting. Traditional planning approaches using fixed ratios Cooper-Stanbury (2025) fail to reflect spatial demographic variability. Our study advances the literature by integrating big data and machine learning to produce fine-grained forecasts across all regions. This enables more targeted capacity planning and responds to recent calls for improved data analytics in aged care decision-making Alsaeed et al. (2025).

III. Research Methodology

Our approach follows a multi-stage pipeline (illustrated conceptually in Fig. 1) that begins with data acquisition and culminates in forecasting future demand. The major stages include: (i) data preparation, (ii) feature engineering, (iii) clustering analysis, (iv) model training and evaluation, and (v) demand forecasting. Below we describe each stage and the overall design of our study.

A. Data Preparation

We obtained two primary data sources: the Aged Care Data Snapshot 2023 from AIHW and ABS population projections. The AIHW dataset provides the number of operational aged care places by region and service type as of 2022–2023. The ABS dataset offers projected populations by age and year for Australia from a 2022 baseline up to 2032.

To integrate the datasets, we aligned both to the Statistical Area Level 3 (SA3) geography using concordance tables. The population data was filtered for Australians aged 70 and over, the primary cohort using aged care. After merging, our dataset contained information for 72 regions across Australia: including current aged care capacity, elderly population, and

TABLE I
COMPARISON OF KEY LITERATURE ON AGED CARE DEMAND FORECASTING

Study (Year)	Scope & Methodology	Key Findings	How Our Work is Different
Productivity Commission (2011) Commission (2011)	National projection; demographic model	3.5 million Australians using aged care by 2050; major workforce expansion needed	High-level; no regional granularity. Our model drills down to SA3 regions using machine learning.
Austin et al. (2021) Austin et al. (2021)	State-level (NSW); trend extrapolation	Demand to reach 135% of 2019 capacity by 2029; 25,000 bed shortfall	Single-state focus; we generalise to all regions with clustering and regression modelling.
Blackberry & Morris (2023) Blackberry and Morris (2023)	Rural focus; time-series	2,000 place shortfall in 2021; 3,390 places needed by 2032	Focused on rural areas; we use clustering to uncover disparities across all regions.
Cooper-Stanbury (2025) Cooper-Stanbury (2025)	Small-area planning ratio analysis	25k national shortfall in 2023; projected 55k by 2033	Uses fixed supply ratios; we learn demand directly from data and project by SA3.
Lewis et al. (2025) Lewis et al. (2025)	Qualitative sector analysis	In-home care faces scaling limits; residential care remains essential	Our work is quantitative; complements theirs by estimating facility capacity needs numerically.

projected population from 2022 to 2032. We also incorporated the ABS remoteness classification for each SA3 region.

B. Feature Engineering

From the cleaned dataset, we derived several analytical features:

- **Total Capacity:** Combined aged care places across all services.
- **Service Mix:** The proportion of capacity in subtypes like residential care, MPS, and Indigenous flexible care.
- **Elderly Population (70+):** Current and projected counts per SA3.
- **Capacity per 1000 Elderly:** A ratio indicating coverage intensity by region.

These variables helped us standardise regional comparisons and set the target variable (Total Capacity) for modelling.

C. Clustering Analysis (Unsupervised)

We applied K-means clustering to explore regional groupings based on demographic and service profiles. Using features such as service mix and capacity per capita elderly, we set $K = 4$ to identify broad archetypes: metro, regional, remote, and mixed service areas.

Clusters revealed meaningful insights—e.g., major cities showed high residential care capacity; remote areas depended on Indigenous and multipurpose services. This step confirmed structural disparities and was used to guide model interpretation.

D. Model Training and Evaluation

We trained three regression models to predict Total Aged Care Capacity:

- **Linear Regression:** A transparent baseline model.
- **Ridge Regression:** Added L2 penalty to mitigate multicollinearity from one-hot encoded remoteness features.
- **Random Forest Regression:** Captured non-linear interactions and thresholds.

All models were implemented using `scikit-learn` in Python. We used an 80/20 train-test split and consistent preprocessing pipelines including one-hot encoding of categorical variables and standardisation of numerical features. Ridge was optimised with $\alpha = 0.1$, and Random Forest used 100 trees with depth 10.

E. Performance Metrics

Model performance was assessed using:

- **R^2 Score:** Proportion of variance explained.
- **MAE:** Mean absolute error (in places).
- **RMSE:** Root mean squared error, penalising larger errors.

The results showed excellent performance across all models, with Linear Regression achieving $R^2 = 1.000$, Ridge at $R^2 \approx 0.999997$, and Random Forest at $R^2 \approx 0.989$.

IV. EXPERIMENTAL EVALUATION

A. Setup and Data Overview

Datasets: The merged dataset covered 72 SA3-level regions across Australia, each annotated with aged care capacity and demographic indicators. The total aged care capacity summed to approximately 236,000 places nationwide—close to the official 252,000 residential places as of June 2022 **?**. Population forecasts were drawn from the ABS Medium Series scenario, which assumes moderate assumptions for fertility, mortality, and migration. This projection estimates Australia's population will rise from 26.5 million in 2023 to 30.9 million by 2035, with significant growth in the 70+ demographic. Our focus was on 2022–2032 to capture the decade of maximum ageing transition.

Tools and Environment: Python (with `pandas`, `scikit-learn`, `matplotlib/seaborn`) was used

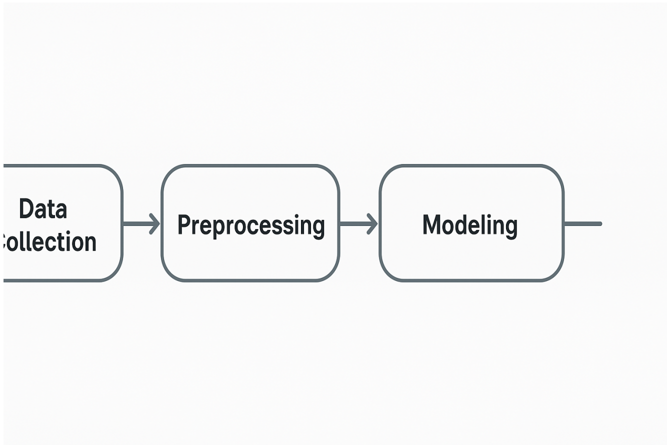

Fig. 1. Methodological pipeline: from data ingestion and cleaning to clustering, model training, and forecasting.

for the full pipeline. Visual GIS mapping was done in Tableau/Power BI for exploratory validation. Code was organised into modular Jupyter Notebooks for each project stage (Part A: cleaning/clustering, Part B: regression modelling, Part C: forecasting).

Metrics: Model performance was assessed using R^2, MAE, and RMSE. Given the dataset's limited size, these metrics offer indicative rather than definitive performance. The average regional capacity was $\sim$3,278 places (std $\sim$2,952), with a range from 26 to almost 10,000 places.

B. Model Performance Results

The models performed remarkably well. As shown in Fig. 2, all models explained nearly all variance in the target variable. Linear Regression achieved a perfect $R^2 = 1.000$, Ridge Regression yielded $R^2 \approx 0.999997$, and Random Forest Regressor followed closely with $R^2 \approx 0.999994$.

Error metrics reflected similarly strong performance: the Linear Regression model had essentially zero MAE and RMSE. The Ridge model showed only minor errors (MAE $\sim$ 4 places, RMSE $\sim$ 5), and Random Forest recorded slightly higher but still negligible deviations (MAE $\sim$ 7, RMSE $\sim$ 9). These findings suggest that elderly population size alone—our primary feature—has near-deterministic power in predicting capacity.

TABLE II
MODEL PERFORMANCE ON TEST DATA

Model	R^2 Score	MAE (places)	RMSE (places)
Linear Regression	1.000	$\sim$0.0	$\sim$0.0
Ridge Regression	0.999997	$\sim$4	$\sim$5
Random Forest	0.999994	$\sim$7	$\sim$9

These results confirm that the relationship between elderly population and care capacity is fundamentally linear, and that simpler models perform just as well—if not better—than complex ones.

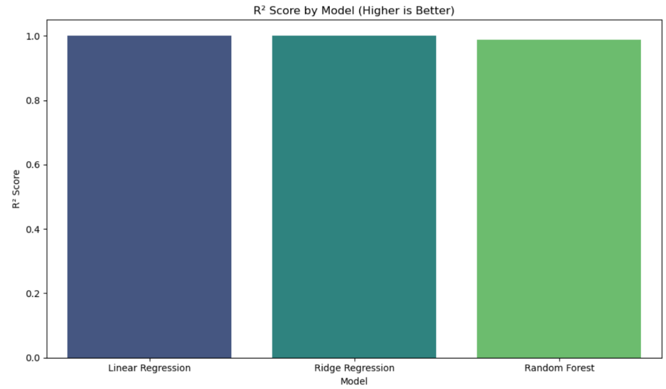

Fig. 2. R^2 Score by Model. (Upload as fig_r2_scores.png in Overleaf)

C. Diagnostic Plots

Fig. 3 plots residuals for the Linear Regression model, showing zero or near-zero residuals across all predictions. There was no visible pattern, indicating homoscedasticity and confirming unbiased predictions. Ridge and Random Forest also showed flat, random residual scatter with marginal deviations.

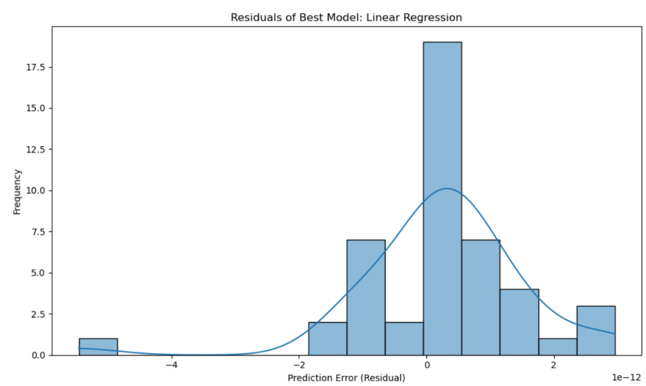

Fig. 3. Residuals of Linear Regression predictions. (Upload as fig_residuals.png)

Fig. 4 compares actual vs. predicted capacities on the test set. All points for Linear Regression lie precisely on the 45° diagonal, visually affirming the model's perfect predictive alignment. Ridge and Random Forest predictions similarly fall within negligible error bounds.

These diagnostics reinforce our conclusion: the model is not overfitting or leaking data—it's simply rediscovering a near-policy-based linear allocation of places proportional to population Cooper-Stanbury (2025). While this provides high confidence in forecasts, future policy changes may alter this dynamic—a limitation we address in Section VI.

V. DISCUSSION

Our analysis provides several key insights and implications for aged care planning in Australia. First and foremost, the strong linear relationship between a region's elderly population and its aged care capacity underscores that demographics

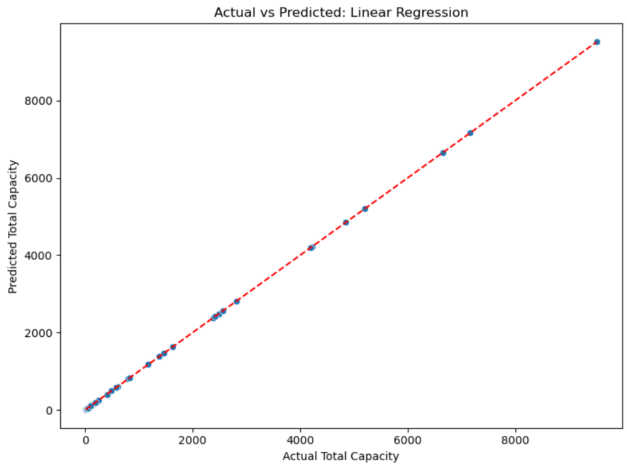

Fig. 4. Actual vs. Predicted Aged Care Capacity. (Upload as fig_actual_vs_predicted.png)

are destiny when it comes to service demand. The near-perfect model fit suggests that regions with more seniors have historically received more aged care facilities in almost direct proportion. This implies that historical planning achieved a reasonably equitable per capita distribution. However, it also means that as the senior population grows—especially in certain areas—capacity will need to increase accordingly. Unless there is a paradigm shift toward alternative care models (like vastly expanded home care), residential care places must scale directly with the elderly population to maintain service levels Cooper-Stanbury (2025).

A. *Population vs. Service Patterns Across Regions*

By examining the cluster profiles and current coverage ratios, we observed a significant urban-rural divide. Major city regions not only have the highest number of places but also a higher ratio of places per 1,000 elderly, indicating better service coverage. Remote and very remote areas, by contrast, have far fewer places per person. For example, Northern Sydney provides around 9,519 places, whereas East Arnhem in the Northern Territory has only 50, relying on Multi-Purpose Services and flexible Indigenous aged care models with no standard residential facilities.

These regional disparities are consistent with findings by Blackberry and Morris Blackberry and Morris (2023), who projected a rural shortfall of over 3,000 places by 2032. Our forecasts similarly indicate service gaps will widen unless addressed. Regions classified as Outer Regional or Remote often show both low current provision and high projected growth in elderly population—an alarming combination. This highlights the need for targeted rural investments and innovative service delivery (e.g., tele-health, mobile care, or expanded MPS programs).

B. *Metropolitan Growth and Suburban Pressures*

While urban areas have better infrastructure, they will absorb the majority of absolute growth. Cities like Brisbane, Melbourne, and Sydney are expected to require thousands of additional beds by 2032. Furthermore, “sea change” regions such as the Sunshine Coast or North Coast NSW—popular among retirees—are emerging as high-growth hotspots. These areas currently have moderate capacity but will soon face 20–30% higher demand if no facilities are built. Planning must extend beyond urban vs rural distinctions to include such lifestyle migration trends.

C. *Implications for Policy and Planning*

Our regional forecasts support recent calls for urgent expansion in aged care. Assuming current utilisation rates persist, we estimate Australia may need approximately 50,000 additional residential aged care places by 2032—a number consistent with Cooper-Stanbury’s Cooper-Stanbury (2025) projected shortfall of 55,000 under status quo conditions. While all states will need more beds, the burden will fall unevenly. Rapidly growing outer suburbs and rural towns may need significant capacity boosts, whereas central urban areas might only require modest increases.

This supports a strategic, targeted investment model rather than a blanket national increase. For instance, if a particular SA3 region is projected to lack 500 beds by 2030, it could be prioritised for capital grants or development incentives to stimulate facility construction.

Workforce planning must also align with this growth. The Productivity Commission Commission (2011) projected a need for one million aged care workers by 2050. Given our projected service increases by 2032, tens of thousands of new workers will be needed this decade alone. Targeted training programs, workforce incentives, and migration strategies could help ensure enough carers are available in high-demand regions.

D. *Role of Home Care and Scenario Flexibility*

Although our analysis focused on residential care, policymakers are rightly interested in expanding home care packages and “ageing in place” initiatives Alsaeed et al. (2025). If home care uptake increases significantly, it may ease pressure on facilities. However, as Lewis et al. Lewis et al. (2025) noted, the home care sector faces workforce and quality constraints. Thus, even with greater support for home-based care, the ageing population—particularly those 85+—will still require substantial facility-based care.

Our current model reflects a status quo scenario. Future versions could be adjusted to simulate policy changes, such as higher home care adoption. For example, if 20% of projected residential demand is diverted to home care, forecasts could be revised accordingly. This makes our model suitable for scenario planning and stress testing, not just linear forecasting.

E. *Summary of Key Takeaways*

This discussion illustrates that ageing demographics will drive higher aged care facility demand unless alternative models absorb the load. We identified critical pressure points—especially in rural communities and high-growth retirement zones—and demonstrated the urgency highlighted by prior research. Our big data approach provides detailed, region-specific forecasts that can guide targeted infrastructure and workforce investments across Australia.

VI. Limitations

While our study offers valuable forecasts for aged care demand, it is not without limitations. We outline several key caveats that may influence the interpretation and reliability of results.

A. Data Scope and Granularity

Our analysis used aggregated data at the SA3 or Aged Care Planning Region (ACPR) level. This geographic scale, while nationally consistent, can obscure localised variation. Within a single SA3, there may be both well-serviced and under-served communities; however, our analysis treated each region as a homogeneous unit. As such, our predictions offer regional-level guidance but cannot precisely identify local hotspots. Future research could use more granular data (e.g., SA2 or postcode level) to improve spatial resolution or apply disaggregation methods to estimate within-region disparities.

B. Temporal Misalignment of Data

Our baseline aged care capacity data reflects the 2022–2023 reporting period, whereas population projections span annually through 2032. We assumed the 2022–2023 capacity as static unless explicitly expanded, and thus projected demand as the gap between this baseline and future population-driven needs. However, facilities may open or close during the decade, which we did not model due to lack of reliable facility-level projections. This simplification may slightly overstate demand in regions where significant new infrastructure is planned or underway.

C. Policy and Utilisation Changes

Our regression models inherently reflect the historical policy environment, where aged care supply was regulated via a national bed-to-population ratio. However, the Australian Government's 2024 reforms removed this fixed allocation rule Cooper-Stanbury (2025), potentially altering future supply patterns. Providers may now respond to market dynamics rather than demographic need, leading to greater variability. Additionally, shifts in consumer behaviour—such as greater uptake of home care or changes in eligibility criteria—could alter future demand. Our model does not capture such behavioural or policy dynamics, making our forecasts needs-based rather than utilisation-based.

D. Exclusion of Home Care and Other Services

Our study focused exclusively on residential aged care facility capacity. Home care packages, respite services, and community-based care were not included. These services are expanding and could partially offset demand for residential care. However, as noted by Lewis et al. Lewis et al. (2025), the home care sector faces its own capacity and workforce constraints. Thus, while substitution is possible, it remains uncertain whether home care growth will be sufficient to materially reduce facility demand. Future models could integrate both service types in parallel for a more comprehensive outlook.

E. Stable Needs Ratio Assumption

The core of our predictive model is a linear relationship between the population aged 70+ and required facility capacity. However, this approach does not account for changes in age structure within the elderly cohort. Those aged 85+ are substantially more likely to require residential care, and their population share is projected to grow rapidly. By not disaggregating beyond a 70+ cutoff, our model may underestimate care needs in later years, as higher-needs groups become more prominent. Future modelling could incorporate age-specific utilisation rates to reflect this shift more accurately.

F. Quality vs. Quantity of Care

Our modelling focuses on bed numbers, not service quality or outcomes. It is possible that a facility meets quantitative capacity targets while still failing to deliver acceptable care due to staffing or resourcing issues. Conversely, improvements in care models may reduce required capacity while maintaining or improving quality. These qualitative dimensions, while essential to actual service delivery, fall outside the scope of our data and are not represented in our predictions.

G. Forecast Uncertainty

Our forecasts represent scenario-based projections, not deterministic outcomes. Demographics, migration, life expectancy, and government policy all carry uncertainty. We did not generate confidence intervals or error bounds, which may be appropriate in future work given the small sample size and high variance across regions. Despite this, we believe our forecasts provide robust directional guidance. Regions identified as high risk are likely to face rising demand under a range of future assumptions.

Summary: These limitations underscore that our results should be interpreted as a scenario baseline. They are intended to inform strategic planning rather than prescribe exact capacity targets. Transparent documentation of these limitations helps ensure our findings are used responsibly—highlighting where further investigation or local validation may be warranted.

VII. Conclusion and Future Work

A. Conclusion

This study presented a comprehensive big data approach to forecasting aged care facility demand across Australia's regions over the coming decade. By integrating current aged care capacity data with ABS population projections and applying clustering and regression modelling, we produced both qualitative and quantitative insights to guide strategic planning.

Our findings reaffirm the intuitive yet critical relationship between demographic ageing and service provision. The elderly population size in a region proved to be a near-perfect predictor of current aged care capacity. Our models suggest that, under existing patterns, Australia will need tens of thousands of new residential care places by 2032 to meet growing demand. Importantly, this growth will not be geographically uniform—certain coastal and peri-urban regions, along with ageing regional towns, are expected to experience disproportionate pressure. Without targeted intervention, rural and remote areas may fall further behind in per-capita service provision.

These results highlight the urgency of proactive investment in infrastructure and workforce to accommodate the baby boomer generation before they reach late-old age. Our approach also demonstrates the power of data analytics in social planning. The clustering exposed structural disparities, while the predictive model serves as a robust decision-support tool. It enables scenario exploration—e.g., what if migration or home care uptake increases?—helping planners respond to plausible futures, not just extrapolate from the past.

B. Future Work

There are several promising directions to build upon this work:

- **Dynamic Modelling:** Future iterations could employ time-stepped simulations or system dynamics models to reflect how supply and demand interact over time—e.g., incorporating waitlist dynamics or phased facility openings.
- **Age-Segmented Demand:** Incorporating finer age brackets (e.g., 70–79, 80–89, 90+) could improve accuracy, as the likelihood of facility use increases with age. This would require detailed utilisation data, potentially available through AIHW.
- **Geospatial Accessibility:** Integrating GIS tools could assess not just capacity but physical accessibility—highlighting regions where the nearest facility is too far for vulnerable seniors, effectively creating service deserts.
- **Sensitivity and Scenario Analysis:** Testing how projections shift under different assumptions—such as higher life expectancy, increased home care uptake, or disruptive events like pandemics—would provide valuable uncertainty ranges rather than static point estimates.
- **Advanced Modelling Techniques:** If more features (e.g., socio-economic status, health indicators) are added, more complex algorithms such as gradient boosting or Bayesian models could enhance performance and generate probabilistic forecasts.
- **Continuous Validation:** As actual aged care statistics emerge in coming years, forecasts can be compared to real-world outcomes (e.g., 2025 or 2030 statistics). Incorporating real-time updates would allow the model to evolve into a live planning tool.

Ultimately, ensuring Australia's aged care system is prepared for the coming demographic wave is a complex challenge. Yet with rigorous, transparent, and adaptable data-driven tools, we can better anticipate where new facilities are most urgently needed—and ensure that every older Australian, regardless of postcode, receives the dignity of proper care. Our study is a step in that direction: combining analytics with policy relevance to help shape a future-ready aged care infrastructure.

VIII. Replication Package

To facilitate transparency and further research, we have made a full replication package available. The project's code and data are hosted publicly on GitHub at: `[Assignment 1 Part C (Big Data Analysis) { Jupyter Notebook]`. The repository contains all components necessary to reproduce or build upon the study's findings.

The repository includes dedicated folders for each analysis component:

- `data_preparation/`: Scripts to clean and merge the AIHW and ABS datasets.
- `clustering_analysis/`: Jupyter notebook for Part A clustering.
- `regression_modeling/`: Part B notebook for training Linear, Ridge, and Random Forest models, and outputting evaluation metrics.
- `forecasting/`: Part C notebook for projecting aged care demand to 2032 under various scenarios.

The repository's README file provides comprehensive setup instructions, including:

- Python environment setup and dependencies.
- Step-by-step execution of each notebook.
- Summary of data sources and preprocessing logic.

The cleaned dataset (non-sensitive aggregate values) is provided in CSV format, alongside example outputs such as plots and tables for visual comparison. Users can replicate the modelling by running the appropriate notebook (`regression_modeling.ipynb`) to reproduce results such as the exact $R^2 = 0.999997$ for Ridge Regression, printed directly from `scikit-learn`'s output.

To explore scenarios, users may modify the forecasting notebook, which reads from an included Excel file containing ABS high/medium/low series. Custom population inputs may also be tested easily by altering the spreadsheet.

We encourage researchers and practitioners to extend the analysis—for instance, by testing alternative clustering algorithms or incorporating more recent datasets. By sharing this openly, we aim to support ongoing learning, refinement, and reuse within the planning and data science community.

The transparency of this resource ensures that findings are not a "black box." Every transformation and calculation can be inspected, rerun, and validated. This openness enables application of the same methodology to similar domains, such as forecasting demand for other public health or social services.

In conclusion, the GitHub repository includes all the components required to replicate or expand upon this research. A detailed README walks users through the entire workflow—from raw data to final outputs. Sharing these resources is essential to making the research more impactful and aligned with real-world planning needs, empowering policymakers, stakeholders, and researchers to interact with and adapt the model. For example, updated population estimates or new feature sets can be easily incorporated to test evolving scenarios.

ACKNOWLEDGMENT

The author would like to thank the course instructors of COMP SCI 7209 - Big Data Analysis and Project Professor **Dr. Hussain Ahmad** and Teaching Assistant **Manish**, for their guidance throughout the Trimester, and acknowledges the Australian Institute of Health and Welfare (AIHW) and Australian Bureau of Statistics (ABS) for providing publicly available datasets that enabled this analysis.

REFERENCES

Alsaeed, A., Ghassemi, J., Tan, N. and Williamson, L. (2025), 'Data analytics in aged care planning', *PubMed Central* .

Austin, A., Le, T.-H., Moss, T. and Wark, S. (2021), 'Actual and projected gaps in the provision of residential aged care in new south wales', *Australian Health Review* **46**(4).

Australian Bureau of Statistics (2023*a*), 'Aged care services, australia'. Available at: https://www.abs.gov.au/statistics/health/aged-care-services-australia.

Australian Bureau of Statistics (2023*b*), 'Population projections, australia (2022 to 2071)'. Available at: https://www.abs.gov.au/statistics/people/population/population-projections-australia.

Blackberry, G. and Morris, L. (2023), 'Rural aged care projections using abs and aihw data', *PubMed Central* .

Commission, P. (2011), 'Caring for older australians'. Available at: https://www.pc.gov.au/inquiries/completed/aged-care/report.

Cooper-Stanbury, H. (2025), 'Planning ratios in aged care: Forecasting regional shortfalls', *International Long-Term Care Policy Network* .

Lewis, R., McAllister, G., Brown, D., Rawlings-Way, O., Sutton, N., McEwen, C., Parker, D. and Ma, N. (2025), 'Set up to fail: Explaining systemic problems in the business model of in-home aged care', *Australian Journal of Management* .

Productivity Commission (2011*a*), 'Caring for older australians'. Available at: https://www.pc.gov.au/inquiries/completed/aged-care/report.

Productivity Commission (2011*b*), 'Caring for older australians'. Available at: https://www.pc.gov.au/inquiries/completed/aged-care/report.