MONASH University

Final Report

Disha Rathod

Report for
Data Intelligence and Insights Department of Monash

27 October 2024

MONASH
BUSINESS
SCHOOL

Department of Econometrics & Business Statistics

(04) 0328 1394
drat0009@student.monash.edu

ABN: 12 377 614 012

AACSB ACCREDITED EQUIS ACCREDITED ASSOCIATION OF MBAS ACCREDITED

1 Introduction

2 Background, Motivation

The project is driven by a clear and a very impactful goal where the Australian government has set a target for universities to ensure that at least 20% of their undergraduate students come from Low Socio-Economic Status (SES) backgrounds. This target aims to promote equality in terms of access to higher education, giving students from disadvantaged communities a fairer chance at achieving a university education. However, achieving this goal is not straightforward, it requires a detailed understanding of the current demographics, regional socioeconomic disparities, and the barriers that students face for which we have used SEIFA methodology.

The primary focus of this project is the state of Victoria, which has regions with diverse socio-economic backgrounds. We aim to analyze the feasibility of meeting the government's target by forecasting the Year 12 graduates population which are the students who are potential university candidates—over the coming years. To do this, we're using data from the Australian Bureau of Statistics (ABS), specifically the 2021 Census, which provides insights into age demographics and socio-economic status across Victoria.

The analysis isn't just about numbers; it's about understanding the real challenges that students from Low SES backgrounds face. For example, some of the barriers include geographical distance to campuses, financial pressures that might force students to work rather than study, and differences in access to educational resources. The project will use R programming to visualize these factors on Victoria's map, highlighting where populations are concentrated and identifying areas where students may have fewer opportunities to access higher education.

One key aspect of this project is understanding how the population and socio-economic landscape will evolve. Using statistical tools, we aim to predict changes in the Year 12 graduates population from 2021 to 2030. By combining this demographic forecasting with SES data, we can identify regions that may need targeted support to meet the enrollment goals.

Beyond the analysis, the project has a broader purpose: to provide actionable insights for Monash University. By identifying patterns and trends, we can inform strategies to make higher education more accessible. This might involve expanding campuses closer to disadvantaged areas, offering free travel passes to students from Low SES regions, or designing flexible academic schedules to accommodate those who need to balance work and study. The project's findings could directly shape university policies and government decisions, ensuring that support is directed where it's most needed.

In summary, this project isn't just about hitting enrollment targets—it's about understanding and overcoming the challenges that disadvantaged students face in accessing education. It's a step towards a fairer and more inclusive education system, where all students, regardless of background, have a chance to succeed.

3 Objectives and Significance

3.1 Key Objectives

Our project centered on understanding and forecasting Year 12-ready populations in Victoria from 2021 to 2030, with a focus on socio-economic disparities. By projecting future student numbers, we aimed to identify which regions would experience growth or decline, allowing Monash University to make informed decisions about resource allocation. These projections help Monash anticipate changes in enrollment demands and prepare accordingly, from expanding campus facilities to focusing recruitment and retention efforts where they are needed most.

To get a clear picture of socio-economic challenges, we used the Socio-Economic Indexes for Areas (SEIFA) methodology, which ranks regions based on various indicators like income, education, and employment. This analysis helped us pinpoint the most disadvantaged areas, highlighting where socio-economic barriers could hinder access to education. Understanding these disparities allows Monash University to target its outreach and support programs more effectively, ensuring resources go to regions with the greatest need.

We combined statistical analysis with geospatial mapping to create a comprehensive view of the educational landscape in Victoria. While statistical data gave us the numbers, the maps allowed us to visualize how these figures played out geographically. This combination not only showed us what was happening but also why certain regions faced greater challenges. It provided a clearer context for Monash, helping them understand the socio-economic dynamics at play and informing where to focus efforts for the greatest impact.

The ultimate goal was to provide Monash University with data-driven recommendations to increase Low SES enrollments. By merging population projections with socio-economic analysis, we highlighted key areas for intervention. This guidance helps the university make strategic decisions about where to direct scholarships, outreach, and academic support to ensure it aligns with their mission to improve access for disadvantaged students. Our approach, grounded in both numbers and real-world context, offers a pathway for Monash to foster a more equitable and inclusive educational environment across Victoria.

3.2 Significance of the Analysis and Methodology

Our project aimed to support equitable access to education by using SEIFA to identify socio-economic disparities in Victoria. This analysis helped Monash University pinpoint areas facing the most significant challenges, ensuring their initiatives are grounded in evidence and targeted for maximum impact. The demographic and socio-economic analysis offered a clear roadmap for strategic resource allocation, whether it's expanding facilities in high-growth areas or providing additional support where student numbers are declining. By combining statistical projections with geospatial mapping, we were able to see the human context behind the data, showing how socio-economic conditions intersect with geography. This comprehensive approach provided Monash with valuable insights to address the real challenges faced by students, making their outreach and support efforts more effective. Overall, the project delivered a data-driven view of Victoria's Year 12 landscape, identifying both areas of need and opportunities for growth. This blend of demographic analysis, SEIFA insights, and visual techniques equips Monash with practical recommendations to boost Low SES enrollment and foster a more inclusive education environment.

4 Methodology

4.1 Introduction to Methodology

The goal of this project is to help Monash University meet the government's target of increasing Low SES student participation (i.e. 20%). Over the past 12 weeks, I was working at Monash's Intelligence and Analytics unit, led by Patrick Leung, has been dedicated to this analysis, using a range of data tools and visualization techniques in R.

Our approach began with collecting and preparing data from the 2021 Australian Bureau of Statistics (ABS) Census, which provided crucial information on the Year 12 graduates population and socio-economic indicators. We used R to clean and merge these datasets, managing the challenges of incomplete and unavailable information. The focus was on ensuring that the data was accurate and reliable for further analysis.

The heart of our methodology was projecting how the Year 12 population in Victoria would change from 2021 to 2030, with a particular emphasis on differences across regions with varying SES profiles. R was instrumental in this, with tools like tidyverse helped us handle data, while sf and ggplot2 allowed us to create detailed maps. These visualizations were crucial for spotting regional disparities and identifying areas that might need more support.

Overall, our methodology was about more than just crunching numbers; it was about using data to tell a story and inform strategies for making higher education more accessible and equitable. This approach is designed to support Monash in planning for a fairer education system, ensuring that every student, regardless of background, has a chance to succeed.

4.2 Data Collection and Preparation

###Data Sources

The backbone of this project was the data sourced from the 2021 Census conducted by the Australian Bureau of Statistics (ABS). This data was essential in understanding the demographic distribution and socio-economic factors across Victoria, specifically focusing on students likely to become Year 12 graduates in the coming years. By examining the age demographics and socio-economic indicators provided by the census, we were able to identify which regions had higher or lower socio-economic advantages. This was critical for categorizing socio-economic status (SES) using SEIFA (Socio-Economic Indexes for Areas) methodology.

We utilized data from three geographical levels—Statistical Area 1 (SA1), Statistical Area 3 (SA3), and Statistical Area 4 (SA4). SA1 provided a detailed view, highlighting localized trends and disparities, while SA3 and SA4 gave us a broader regional perspective. This multi-level approach was crucial for capturing both localized and state-wide patterns, allowing us to identify areas of disadvantage or opportunity accurately.

To streamline the data and focus specifically on Year 12 readiness, we used the ABS Table Builder. This tool allowed us to create custom datasets that concentrated on the age groups relevant to our study and narrowed down the socio-economic factors affecting educational access. Additionally, we incorporated data on university campus locations to understand how accessibility might vary by region, offering insights into the potential impact of distance and proximity on Low SES students.

###Data Preprocessing

Preparing the data for analysis was a critical step, ensuring accuracy and relevance for our projections. Here are the main tasks we undertook:

```
[1] 6336925
```

```
# A tibble: 10 x 2
   Year  Total_Value
   <chr>       <dbl>
 1 2021       70594.
```

Table 1: *Total Values Summarized by Year*

Year	Total_Value
2021	70594.33
2022	70604.33
2023	71845.33
2024	74063.67
2025	76009.33
2026	76796.33
2027	77068.33
2028	77195.00
2029	77943.00
2030	78787.33

```
 2 2022        70604.
 3 2023        71845.
 4 2024        74064.
 5 2025        76009.
 6 2026        76796.
 7 2027        77068.
 8 2028        77195.
 9 2029        77943.
10 2030        78787.
```

```
tibble [190 x 3] (S3: tbl_df/tbl/data.frame)
 $ SA4_NAME_2021: chr [1:190] "Ballarat" "Ballarat" "Ballarat" "Ballarat" ...
 $ Year         : num [1:190] 2021 2022 2023 2024 2025 ...
 $ Total_Value  : num [1:190] 1897 1949 2066 2138 2198 ...
```

```
Rows: 59,280
Columns: 5
$ X2021.Statistical.Area.Level.1..SA1. <dbl> 10102100701, 10102100702, 1010210~
$ State                                <chr> "NSW", "NSW", "NSW", "NSW", "NSW"~
$ Usual.Resident.Population            <int> 305, 301, 471, 522, 423, 290, 416~
$ Score                                <dbl> 984.3059, 1072.3003, 970.2893, 97~
$ Percentile.within.State              <int> 40, 68, 35, 36, 43, 28, 46, 57, 6~
```

```
Rows: 15,014
Columns: 5
```

```
$ SA1reg                    <dbl> 20101100101, 20101100102, 20101100105, 20101~
$ State                     <chr> "VIC", "VIC", "VIC", "VIC", "VIC", "VIC", "V~
$ Usual.Resident.Population <int> 435, 184, 377, 584, 358, 791, 527, 513, 366,~
$ Score                     <dbl> 939.4502, 993.0258, 882.7877, 951.0035, 852.~
$ Percentile.within.State   <int> 22, 40, 9, 25, 5, 36, 15, 69, 61, 48, 54, 38~

Rows: 150,140
Columns: 9
$ SA1reg                    <dbl> 20101100101, 20101100101, 20101100101, 20101~
$ Year                      <chr> "2021", "2022", "2023", "2024", "2025", "202~
$ Value                     <dbl> 9.333333, 7.333333, 7.000000, 6.666667, 6.66~
$ State                     <chr> "VIC", "VIC", "VIC", "VIC", "VIC", "VIC", "V~
$ Usual.Resident.Population <int> 435, 435, 435, 435, 435, 435, 435, 435, 435,~
$ Score                     <dbl> 939.4502, 939.4502, 939.4502, 939.4502, 939.~
$ Percentile.within.State   <int> 22, 22, 22, 22, 22, 22, 22, 22, 22, 22, 40, ~
$ SA4_NAME_2021             <chr> "Ballarat", "Ballarat", "Ballarat", "Ballara~
$ Percentile_Category       <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2,~
```

Reading and Merging Datasets: - We began by importing data from multiple sources using read.csv() and read_excel() from the tidyverse package. Each dataset had distinct structures, with some in CSV format and others in Excel. We cleaned these datasets to maintain consistency. Merging the datasets was an intricate task because they used different identifiers. We performed joins using left_join() from the dplyr package, matching data based on geographical codes such as SA1, SA3, and SA4. This step allowed us to create a unified dataset necessary for further analysis.

Handling Missing Data: - Missing data can skew results, so it was essential to handle any gaps in our datasets. We removed regions with unreliable data using the filter() function and addressed missing values with fill() to maintain data continuity, ensuring logical estimates for gaps. Some datasets contained blank entries for certain regions, particularly those with sparse populations. These were filtered out to avoid inaccuracies in our analysis.

Formatting Age Categories for Projections: - A key element of our analysis was projecting the Year 12-ready population for each year from 2021 to 2030. This involved grouping the age data into three-year segments, such as 17, 18, and 19-year-olds for 2021 projections. We used mutate() and case_when() to create these new variables, enabling us to track changes year-by-year. Rolling averages were calculated using group_by() and summarise() to forecast future demographics, helping smooth out any annual fluctuations.

Data Cleaning and Standardization: - Different datasets often had varying formats, so we cleaned and standardized them to ensure consistency. This included aligning column names, geographic codes, and converting text-based age categories to numeric using as.numeric(). These steps were critical for performing accurate calculations.

Geospatial Data Preparation: - Since much of our analysis relied on visualizing socio-economic trends, we utilized the sf package to handle geospatial data. Shapefiles defining Victoria's Statistical Areas were imported, allowing us to map the data accurately. We performed spatial joins using st_join() to link demographic information with geographic areas. Transformations using st_transform() ensured that all data layers matched spatially, maintaining alignment for accurate visualizations.

Challenges in Data Preparation: - A significant challenge was aligning data collected from different periods and formats. We faced inconsistencies between datasets, particularly in the SES and campus location information, which required careful reconciliation. Using rename() helped standardize inconsistent identifiers, while left_join() enabled precise matching across datasets. Aggregating data from smaller (SA1) to larger (SA3) areas was necessary in regions with incomplete coverage. This aggregation required recalculating averages while preserving regional patterns to ensure our analysis remained accurate.

4.3 Forecasting Methodology

We projected the Year 12-ready population in Victoria from 2021 to 2030 to support Monash University's Low SES enrollment goals using demographic data and R.

To estimate the future Year 12 graduate population, we used age data from the 2021 Census and employed a straightforward averaging technique. The method involved averaging three consecutive age groups to create stable projections. For example, we used the 17, 18, and 19-year-olds to forecast the Year 12-ready population in 2021. For 2022, we adjusted the groups to include 16, 17, and 18-year-olds, continuing this method through to 2030. This rolling average approach smoothed out yearly fluctuations, offering a clearer view of long-term trends without being overly sensitive to annual variations.

The choice to use three-year averages was deliberate, aiming to maintain a balance between detail and simplicity. This technique allowed us to capture natural shifts in age demographics without letting outlier years overly influence the forecasts.

How R Was Used for Forecasting

R played a crucial role in our statistical analysis, offering flexibility and efficiency in data manipulation. We used the `tidyverse` package for data cleaning, structuring, and forecasting. Demographic data was imported with `read.csv()` and refined using `mutate()` and `case_when()`, enabling us to categorize age groups by forecast year easily. This streamlined approach facilitated adjustments to projections and quick insights into demographic trends. Our analysis was regional, focusing on Victoria's Statistical Areas (SA1 and SA3), ensuring forecasts were grounded in local conditions rather than broad state averages. This regional focus allowed us to identify specific areas needing greater support, aligning with Monash University's goals of increasing Low SES student enrollment. The ability to iterate projections and modify parameters seamlessly in R ensured that our analysis could respond to changing data, making it a vital tool in the project's forecasting methodology.

Assumptions and Limitations

Our forecasting methodology relied on several assumptions, which also presented limitations. We assumed stable trends in factors like birth rates and migration, which works well for short-term projections but could miss the impact of sudden socio-economic shifts. The three-year averaging method we employed captured broader patterns effectively, yet it might overlook rapid changes, such as spikes or dips in birth rates. Additionally, our model assumed that students would consistently progress through the education system, without accounting for unexpected dropout rates or shifts in educational policy that could alter projections. The accuracy of our forecasts was heavily dependent on the quality of the 2021 Census data; any inaccuracies in this source could influence our results. Despite these constraints, using R allowed for flexibility in the analysis, enabling real-time data adjustments that made our forecasts more adaptable to emerging trends, while maintaining transparency in the approach.

Geospatial Data Preparation

Mapping socio-economic trends was essential to understanding Victoria's educational landscape, highlighting disparities in population and socio-economic conditions. Using the `sf` package, we conducted geospatial analysis by importing shapefiles that defined Victoria's Statistical Areas—SA1 for detailed regions and SA3 for broader areas. This enabled us to overlay demographic data onto geographic boundaries accurately. Spatial joins using `st_join()` were crucial, linking demographic data to specific locations for precise visualization of trends. To maintain consistency across datasets, we standardized coordinate systems with `st_transform()`, ensuring accurate alignment. This combination of demographic and spatial data provided a visual representation of regional disparities, making it easier to pinpoint areas needing additional support. The geospatial insights revealed how socio-economic factors varied across Victoria, offering a comprehensive understanding of the areas most in need of educational resources.

Creating Maps and Visualizations

Using the merged spatial data, we utilized `ggplot2` to create detailed visualizations that highlighted key educational trends in Victoria. These included maps showing socio-economic conditions, population densities, and proximity to university campuses. Mapping Socio-Economic Status (SES) categories revealed areas with greater barriers to accessing education, guiding targeted interventions. Additionally, demographic density overlays pinpointed regions with high Year 12-ready populations, identifying potential focus areas for boosting university enrollment. These visualizations provided a comprehensive view of the challenges faced by Low SES students and helped inform strategies to make education more accessible.

4.4 Data Analysis and Diagnostics

Our analysis of Year 12 population trends in Victoria was centered around understanding socio-economic differences, specifically to aid Monash University's strategy for increasing Low SES enrollments. Using R, we processed and summarized data to identify key patterns, ensuring data accuracy and reliability throughout.

To get a clear picture, we used `group_by()` in R to segment data by regions and socio-economic status (SES), allowing us to highlight areas with higher concentrations of Low SES students. This information was crucial for pinpointing regions that might need additional outreach and support. To further clarify these patterns, we created SES categories by dividing regions into percentile groups based on their socio-economic rankings. This approach helped visualize the disparities more effectively, showing how socio-economic factors impact access to education.

To ensure the accuracy of our findings, we carried out validation checks by cross-referencing with historical census records and comparing trends across different levels, such as SA1 (local) and SA3 (regional). This step was vital to confirm that local observations were consistent with broader state-wide trends. However, our analysis did have limitations. We focused only on Victoria, which meant that comparisons with other states weren't included. Additionally, we assumed that demographic trends would remain stable, and our data heavily relied on the quality of the 2021 Census, which may have gaps.

Despite these limitations, we continuously adjusted and validated our forecasts to provide Monash University with reliable insights, ensuring their strategies for supporting Low SES students are well-informed and effective in promoting a more equitable educational environment.

4.5 Skills, Challenges, and Innovative Approaches

This project required the application of a broad set of technical skills to effectively manage, analyze, and visualize socio-economic and demographic data for Victoria. The goal was to support Monash University in increasing Low SES enrollment by understanding Year 12 population trends and socio-economic disparities across the state. Using R as our primary tool, we were able to draw valuable insights from the data.

A significant portion of the project involved data wrangling, a critical step in preparing raw datasets for analysis. The data from the Australian Bureau of Statistics (ABS) included various demographic and socio-economic indicators that were not immediately suitable for analysis. Using R's `tidyverse` package, we undertook a meticulous process of cleaning and standardizing the data. This included converting text-based data to numerical formats, aligning regional identifiers, and addressing inconsistencies across datasets. One of the biggest challenges was merging multiple data sources, each structured differently. For example, combining demographic data, socio-economic indicators, and geographic information required careful handling of format discrepancies. We relied heavily on R's merging tools, such as `left_join()` and `inner_join()`, to integrate diverse datasets seamlessly. This effort ensured that our analysis had a solid foundation, free from data inconsistencies.

Understanding the regional differences in Victoria required a detailed geospatial analysis. We used R's `sf` package, which specializes in handling spatial data, to manage the geographical aspects of the project. The analysis involved working with Statistical Areas—SA1 for detailed local data and SA3 for broader regions—ensuring our findings were both specific and comprehensive. By importing shapefiles that outlined these areas, we were able to overlay demographic data onto geographic boundaries, visualizing socio-economic trends across the state. One challenge we faced was aligning datasets with varying coordinate systems. To ensure accuracy, we utilized functions like `st_transform()` to standardize coordinate references, enabling a consistent mapping process. Additionally, we conducted spatial joins with `st_join()` to precisely link demographic data with geographic regions, ensuring that each data point accurately corresponded to its location.

Another key aspect of our analysis was forecasting the Year 12-ready population from 2021 to 2030. This step was crucial for understanding future trends and anticipating areas where Monash University might need to focus its resources. We utilized R's data manipulation capabilities, particularly the `mutate()` and `case_when()` functions, to structure age groups and project them into future years. These tools allowed us to flexibly adjust age groups year by year and average across three age groups to create reliable projections. The challenge here was accounting for potential shifts in demographic patterns, such as changes in birth rates or migration trends. To manage this, we frequently validated our projections by cross-referencing with existing data, refining our models to account for any emerging nuances.

Presenting complex data in a clear and accessible way was an essential part of our project. Using R's `ggplot2` package, we created a variety of visualizations that combined demographic, socio-economic, and geographic data. These visualizations included bar charts, line graphs, and detailed maps that illustrated key trends. For example, maps showing socio-economic status (SES) categories highlighted regions with the greatest educational challenges, while demographic density overlays pinpointed areas with high Year 12-ready populations. These visual tools made it easier to communicate our findings, helping Monash University to quickly identify focus areas for their Low SES strategy.

Integrating multiple datasets with diverse structures and scales was one of the most significant challenges we encountered. We had to standardize region names, formats, and coordinate systems to ensure that all data sources aligned properly. This often required manual adjustments to reconcile discrepancies. In rural areas, where data coverage was sometimes sparse, we aggregated information from smaller SA1 areas to broader SA3 regions to maintain a balance between specificity and reliability. The forecasting component also posed challenges, as it required stable demographic trends to make accurate predictions. We continuously validated our projections, refining them to account for local variations and the inherent uncertainty in long-term forecasts. Balancing the simplicity of our models with the complexity of the data was a constant effort, ensuring that our insights remained both realistic and useful.

Our analysis incorporated several innovative approaches that enhanced the clarity and effectiveness of our findings. One novel aspect of our methodology was the use of percentile categories to represent SES regions. By categorizing Victoria's regions into percentile groups, we were able to standardize how we visualized socio-economic disparities. This approach made it easier to compare different areas, highlighting regions that were more disadvantaged and those that were relatively advantaged. It provided a clear, intuitive way to understand the socio-economic landscape, allowing Monash University to focus on areas that needed the most support.

Another unique strength of the project was the combination of statistical analysis with geospatial visualization. While statistical tools provided the numerical insights, mapping these data points gave them real-world context. This dual approach revealed how socio-economic and demographic factors interacted across Victoria, offering a comprehensive view of educational challenges. The ability to visualize patterns geographically added depth to our analysis, making it easier to communicate complex trends.

The project's methodology was also designed to be adaptive, using R's flexible tools to accommodate new data and unexpected trends. Functions like `mutate()` and `case_when()` allowed us to make quick adjustments and respond to changing variables. This adaptability ensured that our analysis remained relevant, providing Monash with the most up-to-date insights for supporting Low SES students. By building flexibility into our analysis, we were able to refine our recommendations continuously, aligning them with Monash's evolving needs.

Overall, the project's approach combined rigorous technical analysis with innovative visualization techniques to deliver actionable insights, guiding Monash University's strategy for fostering a more inclusive educational environment across Victoria.

4.6 Discussion of Limitations and Data Issues

While our analysis provided valuable insights into Year 12 population trends and socio-economic disparities across Victoria, it was not without challenges. This section explores some of the key limitations we encountered with the data and the potential improvements for future analysis.

Data Limitations

One of the main limitations of our analysis was the lack of detailed socio-economic data. While SEIFA rankings gave a helpful overview, they grouped regions based on factors that might have hid local differences. For example, a low SES area might still have some wealthier pockets that the data didn't show. Additionally, the SES data relied on the 2021 Census, offering a snapshot of socio-economic conditions at that time. However, these conditions can shift rapidly due to changes in local economies, policies, or unexpected events like pandemics, meaning our analysis might not reflect recent socio-economic changes that could impact Year 12 graduates. The forecasting methods also introduced limitations. Projections assumed stable demographic trends over the decade, which may not hold if migration patterns shift, government policies change, or socio-economic landscapes transform. By averaging age groups to estimate future Year 12 populations, we captured broader trends but risked missing sudden local fluctuations. Integrating multiple datasets also posed challenges, with each source having different structures and formats. Standardizing data required extensive cleaning, especially when region names didn't match perfectly. Handling missing data, often aggregating smaller regions (SA1) into larger ones (SA3), sacrificing some detail for consistency.

4.7 Software and Tools

Our analysis relied heavily on a suite of software, packages, and libraries designed to handle data cleaning, analysis, and visualization efficiently. Here's a breakdown of the key tools used:

Studio:

- We used RStudio, a powerful integrated development environment (IDE) for R, as our primary tool for coding, data analysis, and visualization. RStudio provided a user-friendly interface to manage the project's complex datasets, execute scripts, and visualize results seamlessly.

R Packages:

tidyverse: This collection of R packages was fundamental to our data wrangling and manipulation efforts. Packages like `dplyr`, `tibble`, and `tidyr` were used extensively to clean, format, and merge data. **sf**: For handling and visualizing geospatial data, we used the `sf` package, which enabled us to manage shapefiles and perform spatial joins. This package was crucial for creating accurate maps that displayed socio-economic trends across Victoria. **ggplot2**: A core part of the `tidyverse`, `ggplot2` was essential for creating data visualizations. We used it to generate line graphs, bar charts, and detailed maps that captured population projections and SES variations over time. **zoo**: This package was used to manage and analyze time series data, which was particularly useful for projecting Year 12-ready populations over the 2021-2030 period. **readxl**: To import and process Excel files containing demographic and socio-economic data, we relied on the `readxl` package, which facilitated seamless integration of external datasets.

These tools and packages were critical to our workflow, providing the functionality and flexibility needed to handle diverse datasets and create meaningful visualizations that guided Monash University's strategy for supporting Low SES students.

5 Results and Discussion

6 Future work