MONASH University

Forecasting Year 12 Graduates for Monash's Low SES Enrollment Targets

Disha Rathod

Report for
Intelligence and Insights Department of Monash

28 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

Over the past 12 weeks, I have been deeply involved in an analytical journey at Monash University's Intelligence and Insights Committee, where I worked on a project that holds the potential to reshape access to higher education. With a focus on Victoria, my task was to forecast the Year 12 graduate population from diverse socio-economic backgrounds, especially targeting low SES regions. Using the Australian Bureau of Statistics (ABS) data and methodologies like SEIFA (Socio-Economic Indexes for Areas), I developed projections to understand which areas are likely to see significant growth or decline in university-ready students. This data-driven approach wasn't just about numbers, it was about visualizing the real-world barriers and opportunities that shape educational futures.

The implications of this work are far-reaching. For the Victorian Tertiary Admissions Centre (VTAC), which plays a pivotal role in guiding students through the university admissions process, these forecasts provide a clear, data-backed understanding of where efforts are needed most. For more information about VTAC, please refer to this - Victorian Tertiary Admissions Centre (n.d.).

On the university's end, my analysis offers crucial insights for Monash's leadership. The Australian government has set an ambitious target: 20% of university enrollments must come from low SES regions. My forecasts will equip Monash's executives with the data they need to make informed decisions—whether that means investing in outreach programs, enhancing financial aid, or even expanding campus facilities to new areas. The goal is clear: to transform data into strategic actions that create a more efficient and equitable academic environment. By combining detailed demographic projections with insights into socio-economic factors, this project lays the groundwork for Monash University to not only meet the government's target but to surpass it, ensuring that higher education in Victoria is accessible to every student, regardless of their background.

2 Background, Motivation

The motivation for this project stems from a critical objective set by the Australian government, which mandates that universities ensure at least 20% of their undergraduate enrollments come from Low Socio-Economic Status (SES) backgrounds. This initiative aims to achieve equality in access to higher education, allowing students from underprivileged communities to have a fair shot at pursuing a university education. However, reaching this target is complex, as it requires an in-depth understanding of the socio-economic landscape, regional disparities, and the distinct challenges faced by students in lower SES categories. To achieve this, the project leverages the Socio-Economic Indexes for Areas (SEIFA) methodology, as outlined by the Australian Bureau of Statistics (ABS) through their

SEIFA 2021 data Australian Bureau of Statistics (2021), which assesses socio-economic factors such as income, education, and employment across Victoria's diverse regions.

The analysis is focused on Victoria, a state characterized by a variety of socio-economic backgrounds, which makes it an ideal case study for understanding the challenges of meeting low SES enrollment targets. By projecting the population of Year 12 graduates—students who represent potential university candidates—over the coming decade, the study aims to provide data-driven insights. The 2021 ABS Census data, a crucial source for understanding age demographics and socio-economic distribution, serves as the foundation for these forecasts. This is not just an exercise in number-crunching; it is an exploration of the barriers that low SES students face, such as distance to educational institutions, financial constraints that compel them to seek work rather than pursue higher education, and unequal access to essential academic resources.

A major component of the study involves predicting shifts in the Year 12 graduate population from 2021 to 2030. This demographic forecasting, coupled with socio-economic data, will help to pinpoint regions that require focused intervention to meet the university enrollment goals. Using R programming, these insights are visualized on maps of Victoria, illustrating both where populations are concentrated and which areas may be underserved in terms of educational opportunity.

The project's ambition goes beyond simple data analysis. The findings are intended to serve as a strategic guide for Monash University, highlighting where efforts to boost accessibility should be concentrated. This could lead to actionable initiatives such as building new campuses in underserved areas, providing transportation support for students from low SES backgrounds, or creating flexible academic schedules that accommodate students balancing work and study commitments. These insights are not only beneficial for the university's strategic planning but can also influence governmental policies, ensuring that resources are allocated to where they are needed most.

In essence, this project is not solely about meeting quantitative targets but about gaining a deep understanding of the real obstacles faced by disadvantaged students. It aims to dismantle these barriers, paving the way for a more inclusive and equitable educational environment. By addressing these challenges, the project contributes to a vision of higher education where all students, regardless of socio-economic status, have the opportunity to achieve their academic and career goals.

3 Objectives and Significance

3.1 Key Objectives

The primary objective of this project was to forecast the Year 12-ready populations in Victoria from 2021 to 2030, with a distinct emphasis on identifying and understanding socio-economic disparities. By projecting these student numbers, we aimed to determine which regions would see increases or decreases in university-ready populations, allowing Monash University to plan effectively for future enrollment demands. This foresight helps Monash anticipate changes and adjust its strategies, whether that involves expanding campus infrastructure, enhancing recruitment efforts, or developing targeted retention initiatives to support specific regions.

To achieve a nuanced understanding of socio-economic challenges, we utilized the Socio-Economic Indexes for Areas (SEIFA) methodology developed by the Australian Bureau of Statistics (ABS), as outlined in SEIFA 2021 Australian Bureau of Statistics (2021). This method ranks regions based on factors such as income, education, and employment, enabling us to identify the most disadvantaged areas in Victoria. These insights allowed Monash University to focus its outreach and support programs on regions with the highest need, ensuring that resources are directed where they can have the greatest impact.

Our analysis integrated statistical modeling with geospatial mapping, providing a dual perspective on Victoria's educational landscape. While statistical projections offered quantitative data, mapping visualized these numbers within their geographic and socio-economic contexts. This combination not only identified the extent of disparities but also clarified why certain regions faced greater challenges. This contextual understanding was crucial for Monash, helping them comprehend the socio-economic dynamics influencing educational access and allowing them to prioritize their efforts accordingly.

The overarching goal was to deliver actionable, data-driven recommendations to Monash University, guiding their strategy for increasing Low SES enrollments. By blending demographic forecasting with socio-economic insights, we pinpointed regions most in need of support. These recommendations are intended to steer Monash's decision-making in allocating scholarships, directing outreach programs, and enhancing academic support for students from disadvantages regions. Our evidence-based approach provides a concrete path for Monash to create a more inclusive and accessible educational environment across Victoria.

3.2 Significance of the Analysis and Methodology

This project holds significant value in supporting equitable access to education by using SEIFA to expose socio-economic disparities within Victoria. By pinpointing regions with substantial socio-economic challenges, Monash University can ground its initiatives in solid evidence, targeting efforts for the greatest possible impact. Our approach offered a strategic guide for the allocation of resources, from expanding facilities in areas with anticipated growth to concentrating support in regions where Year 12 graduate populations are predicted to rise.

One of the key strengths of our methodology lies in its ability to present data in both statistical and visual forms. Statistical projections provided the necessary quantitative analysis, but by pairing these numbers with geospatial mapping, we could illustrate the broader context and human stories behind the figures. This visual representation clarified how socio-economic factors intersect with geography, revealing patterns that pure statistics might overlook. The comprehensive view provided Monash with a deeper understanding of the real-world challenges faced by students, making their outreach efforts more targeted and impactful.

The project's innovative combination of demographic forecasting, SEIFA insights, and visual analytics resulted in a thorough understanding of Victoria's Year 12 landscape. By blending multiple data sources, we created a nuanced analysis that not only highlighted areas of disadvantage but also identified opportunities for growth and investment. This multifaceted approach equips Monash University with practical, evidence-based recommendations to enhance Low SES enrollment, advancing their mission to foster a more equitable education system. Overall, the project delivered a robust, data-driven foundation for making informed decisions, helping Monash lead the way in creating an inclusive and supportive educational environment for all students.

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 Insights 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. To learn more about this Table Builder, you can visit - Australian Bureau of Statistics (2024).

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:

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. Total population was calculated for Victoria for the year 2021 which came out to be 6336925 which was very close to the population reported by ABS **refer1**. 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.

Table 1: *Year 12 Graduate Population Projection*

Year	Population
2021	70594.33
2022	70604.33
2023	71845.33
2024	74063.67
2025	76009.33
2026	76796.33

In Table 1, we can see the projected values for Year 12 graduate population that will be ready every year and this table will also help us build the trend line.

Table 2: *Summarized Year 12 Population values by SA4 Region and Year*

Region	Year	Population
Ballarat	2021	1897.000
Ballarat	2022	1949.333
Ballarat	2023	2065.667
Ballarat	2024	2138.000
Ballarat	2025	2197.667
Ballarat	2026	2236.667

In Table 2, I have taken the projected Year 12 population values every year as Table 1 and divided them by SA4 regions each, as this makes our analysis easier in the further part of the project.

```
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~
```

This dataset gives us the SES status defining variables in detail for each SA1 region in whole of Australia.

```
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~
```

Here, we have filtered only the SA1 regions in Victoria, as we are only focusing on VIC for our project.

```
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,~
```

This is the dataset that we have created by joining the dataset which had the population values of Year 12 graduates in Victoria with the SES defining data.

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

5.1 The line graph for the projections made for Year 12 population in Victoria

As shown in Figure 1, the projected Year 12 graduate population in Victoria from 2021 to 2030, the x-axis represents the years, ranging from 2021 to 2030, while the y-axis indicates the total projected population of Year 12 graduates.

Analysis of the Graph: The trend line, marked in blue, showcases a general upward trend in the Year 12 population over the decade. From 2021 to 2023, there is a relatively modest increase in the

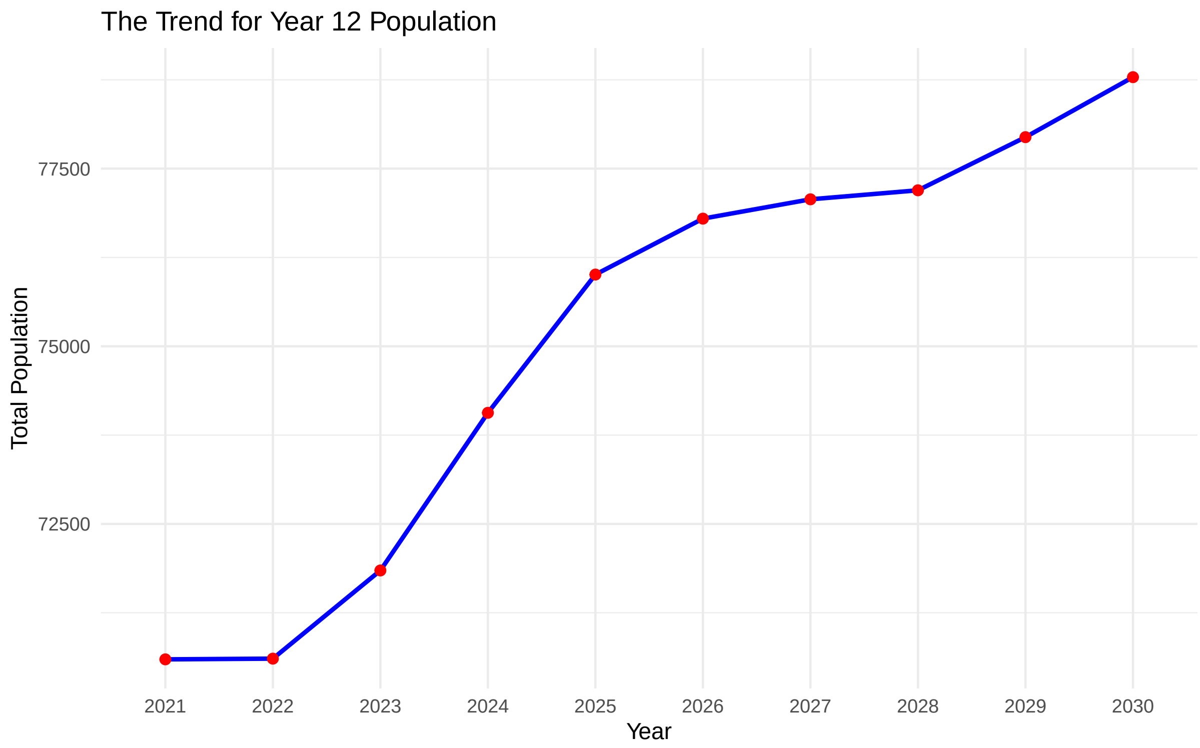

Figure 1: *The Trend for Year 12 Population*

population, suggesting steady growth. After 2025, the graph shows a continued but steadier incline, suggesting sustained growth in the Year 12 population until 2030, mostly at a slightly slower rate. The red dots on the graph indicate specific data points for each year, emphasizing the consistency of the growth pattern over time. The visual clearly points out that by 2030, the Year 12 graduate population is expected to be substantially higher than the preceeding years, reflecting a positive trend in education completion rates.

Relation to Project Objectives: This projection is directly tied to the core objectives of the project, which aims to assess and project the Year 12 population trends in Victoria, with a particular focus on students from Low Socio-Economic Status (SES) backgrounds. The increasing trend seen in the graph suggests a growth of potential university candidates, making it crucial to understand which regions might face challenges in accessing higher education due to socio-economic barriers.

The projection helps inform strategic planning, allowing targeted outreach efforts, resource allocation, and policy adjustments to ensure that educational opportunities are equitably distributed, particularly for disadvantaged groups. This aligns with the goal of enhancing access for Low SES students and meeting the government's target for increased university participation from these backgrounds.

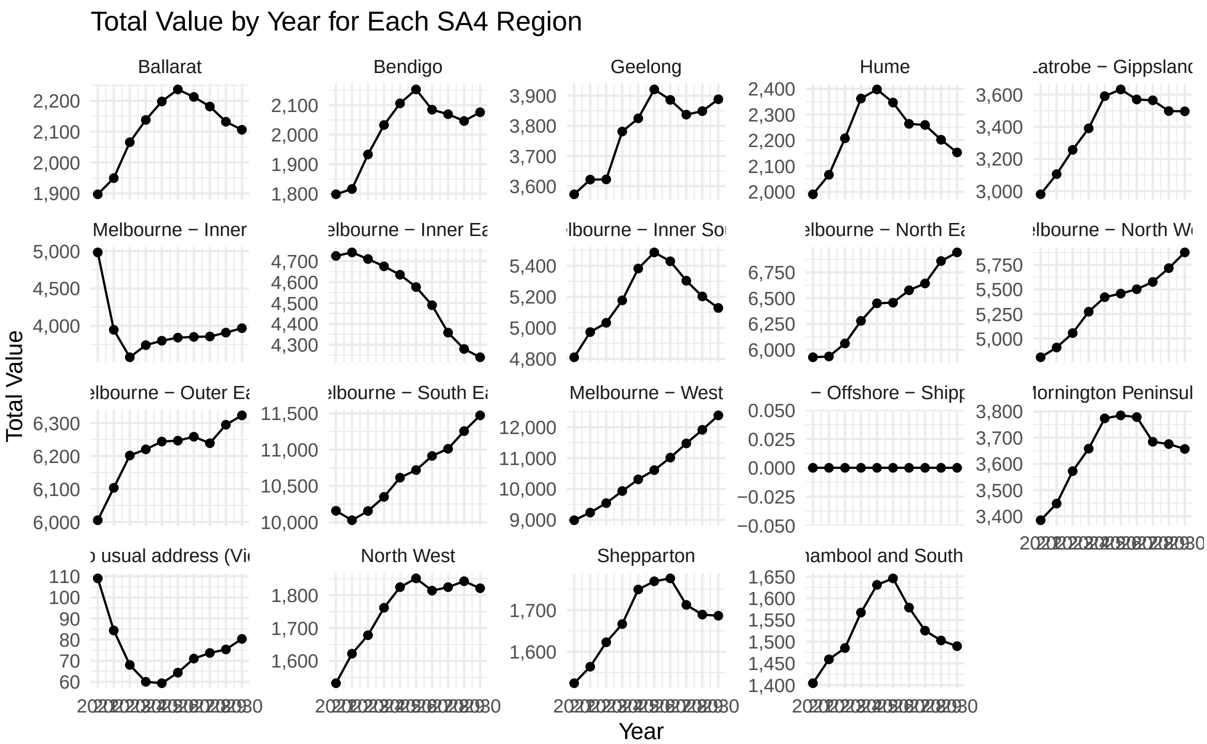

Figure 2: *Total Value by Year for Each SA4 Region in Victoria (2021-2030)*

5.2 The projection of the population values forecasted for each SA4 region in Victoria

The Figure 2 displays the projected Year 12 graduate population from 2021 to 2030 across various SA4 regions in Victoria. Each subplot represents a different region, allowing us to visualize how the Year 12-ready population is expected to evolve over the decade. The x-axis indicates the years from 2021 to 2030, while the y-axis shows the projected population values. The solid line in each subplot connects the data points year by year, providing a clear trend for each region.

Analysis of the Graph: The projections vary significantly across regions, reflecting diverse local population trends. For instance, regions like Melbourne - South East and Melbourne - West show a consistent upward trend, indicating steady growth in the Year 12 population(which are the areas closer to Monash campuses). These areas are likely to see an increasing number of graduates over the next decade. In contrast, regions such as Melbourne - Inner East and Shepparton exhibit a peak followed by a decline, suggesting that the Year 12 population might decrease after hitting a peak around the mid-2020s.

Relation to Project Objectives: The Figure 2 directly aligns with the project’s objectives of understanding Year 12 demographic trends in Victoria. By breaking down projections by SA4 regions, the graph enables Monash University to pinpoint which areas might face higher educational demand and where interventions may be necessary. For regions showing strong growth, strategic planning can involve increasing outreach, resources, or infrastructure to accommodate more students. Conversely, regions showing potential declines may benefit from retention strategies or targeted programs to ensure students remain engaged and are encouraged to complete Year 12.

5.3 Trend for each Percentile Category in every SA4 region in Victoria (Red = LowSES)

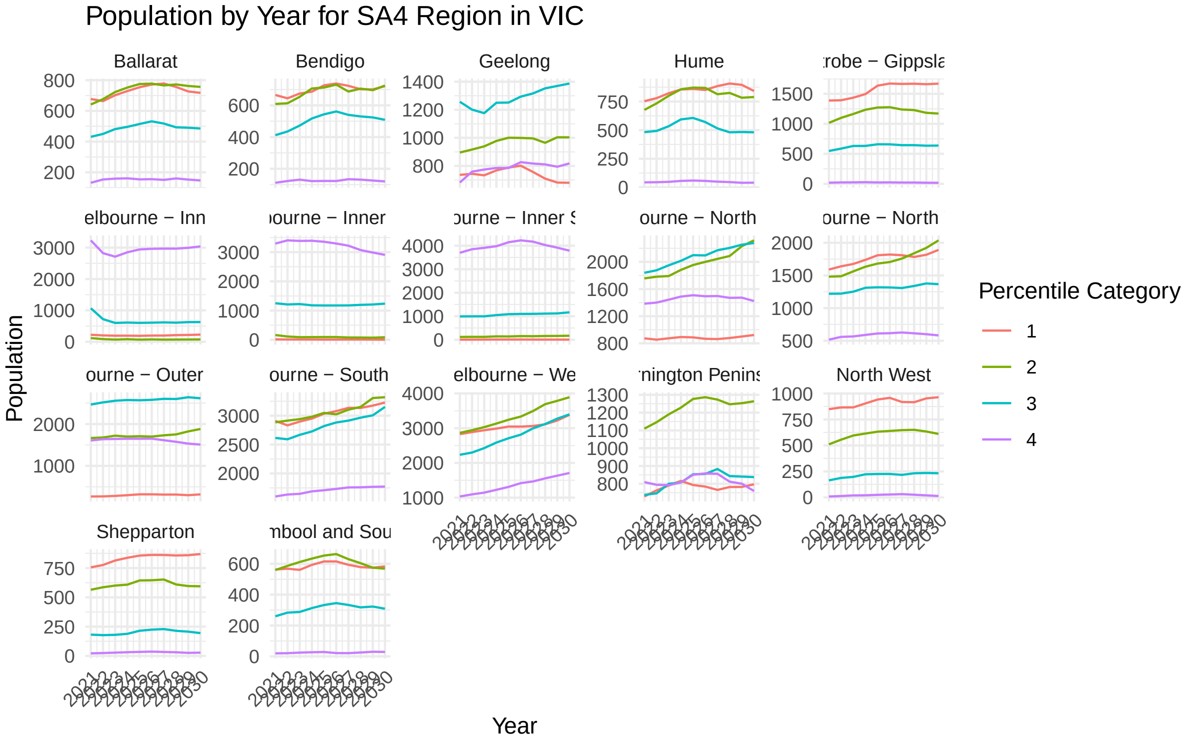

Figure 3: *Yearly Population Trends by SA4 Region and Percentile Category in Victoria (2021-2030)*

The Figure 3 illustrates the projected Year 12 population trends across various SA4 regions in Victoria, broken down by socio-economic percentile categories. Each line color corresponds to a different SES group, with the red line representing the lowest socio-economic status (most disadvantaged) and the purple line indicating the highest socio-economic status (most advantaged). Green and blue lines depict the middle-low and middle-high SES groups, respectively.

The Melbourne North-west (closer to our campuses) and North West have an upwqard trend line for Low SES population for year 12, that can help us make further decisions.

5.4 Tha SES map of 2021 (Socio Economic status of regions in central part of Victoria for SA3)

The Figure 4 illustrates socio-economic disparities across Victoria, with darker shades indicating advantaged regions and lighter shades highlighting disadvantaged areas. Monash University's main campuses, such as Caulfield and Clayton, are surrounded by darker, advantaged areas, suggesting that students nearby likely have better access to resources and educational opportunities. In contrast,

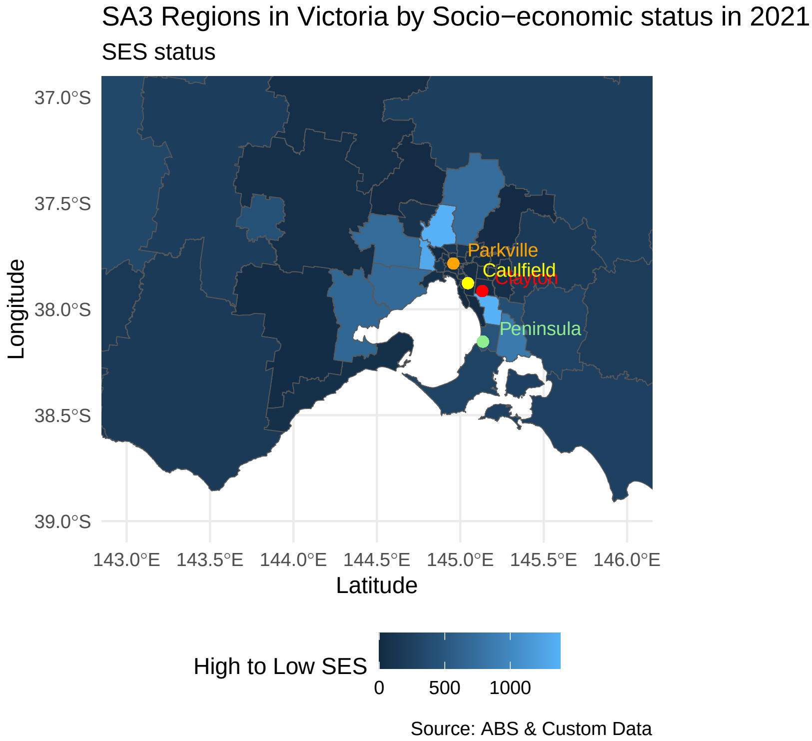

Figure 4: *Socio-Economic Status Across SA3 Regions in Victoria in 2021*

the lighter-shaded regions, particularly in the northern, western and outer areas, represent socio-economically disadvantaged regions where students face more barriers to higher education, such as limited resources and financial challenges. This visual insight suggests that Monash could focus outreach, scholarships, and support programs in these lighter regions to increase participation from low SES backgrounds. Additionally, strategic planning could involve expanding facilities/campuses or forming partnerships in low SES areas to reduce geographical and socio-economic barriers, aligning with Monash's objective of fostering a more equitable education system across Victoria.

5.5 Map of SA3 regions in central victoria (zoomed-in) with the year 12 population projections for 2030

This Figure 5 displays the projected distribution of Year 12 graduates across Victoria's SA3 regions for 2030, with shading indicating population concentrations. Darker areas represent lower projected Year 12 populations, while lighter regions show higher numbers. Notably, the lighter shades, concentrated around Melbourne's metropolitan regions near Monash campuses such as Clayton and Peninsula, suggest a higher influx of potential university applicants in these areas. In contrast, the darker shaded areas, particularly the Melbourne CBD predict lower Year 12 populations.

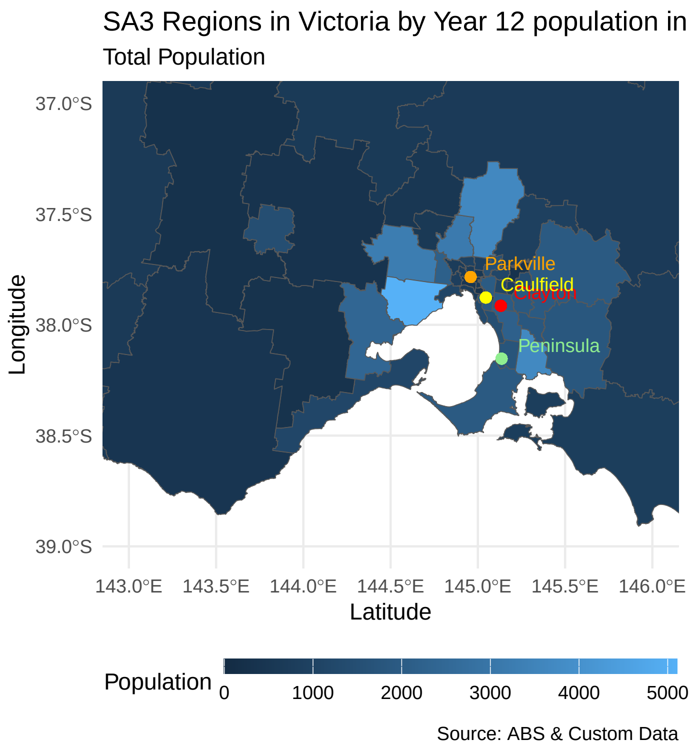

Figure 5: *Projected Year 12 Population in SA3 Regions of Victoria for 2030*

6 Future work

The study provides Monash University with a robust foundation to strategically expand its reach and better meet the government's Low SES enrollment targets. By using demographic forecasts and socio-economic data, Monash can make informed decisions on where to focus recruitment efforts, allocate resources, and develop targeted outreach programs. Future work could involve integrating this analysis with confidential VTAC data to gain deeper insights into student application trends and outcomes, allowing for a more refined approach to student support. Additionally, expanding geospatial analysis and engaging in scenario planning could help Monash explore how changes in policy or new initiatives might impact student demographics. This combined data-driven approach will empower Monash to optimize resource distribution, anticipate enrollment demands, and ultimately enhance its decision-making for future growth.

7 Conclusion

This project set out to forecast the Year 12 graduate population in Victoria from 2021 to 2030, focusing on the socio-economic factors of the regions. Our aim was to provide Monash University with the insights needed to enhance participation for students from low socio-economic backgrounds, in line with the Australian government's target of having 20% of university enrollments come from these regions. The findings from the analysis offered a comprehensive view of demographic trends across Victoria, highlighting opportunities and challenges for creating a more equitable education system.

One of the most striking results was the steady increase in the Year 12 graduate population across Victoria, particularly in regions close to Monash's key campuses. Areas like Melbourne's South-East and Western suburbs are projected to see significant growth in university-ready students in or around 2030. These are critical for Monash, as they guide decisions about where to direct resources and how to support regions that might otherwise face a decline in university candidates. By leveraging SEIFA data for the SES status of the regions, the analysis pinpointed regions with the highest concentrations of Low SES students, suggesting a clear strategy for Monash to focus on outreach, scholarships, and academic support. This targeted approach will be crucial for fostering an inclusive academic environment.

The study also has significant policy implications. Monash University can use these findings to make informed decisions on campus expansion, outreach efforts, and the allocation of scholarships. Understanding demographic trends and the socio-economic challenges associated with them allows for strategic planning that is both evidence-based and effective. This data can help Monash develop initiatives to reach low SES regions and boost Low SES enrollments, contributing to a fairer and more equitable education system. Expanding campus facilities, partnering with local schools, and offering flexible learning options are some practical ways to address the educational needs of these disadvantaged regions.

The study, however, is not without limitations. The projections depend on stable demographic trends, assuming that factors like birth rates and migration will remain consistent over the forecast period. In reality, unexpected shifts—such as economic changes or global events—could alter these predictions. Moreover, the socio-economic data was based on the 2021 Census, which might not fully capture recent changes in the socio-economic landscape. Updating the analysis with more recent and localized data in the future would offer a more accurate and current understanding. Additionally, this study focused solely on Victoria, so expanding the analysis to include comparisons with other states could provide a broader perspective on educational inequalities across Australia.

Despite these limitations, the project offers a solid foundation for Monash University to understand and address educational barriers. By forecasting Year 12 graduate numbers and mapping socio-economic factors, it provides a data-driven framework for meeting government targets and supporting students from disadvantaged backgrounds. The study opens the door for further research, including more detailed analyses that integrate additional data sources like VTAC application trends, and broader geographical comparisons. This kind of continued analysis will help refine and strengthen

Monash's strategies for supporting an inclusive educational environment, ensuring that students from all backgrounds have the opportunity to succeed.

The work done here lays the groundwork for a fairer future in education, one where strategic, data-informed decisions can directly impact the lives of students from all socio-economic backgrounds. With a focus on evidence-based policy and practical interventions, Monash is well-positioned to lead the way in fostering greater access to higher education across Victoria and beyond.

The git repo can be accessed using this link

References

Australian Bureau of Statistics (2021). *Socio-Economic Indexes for Areas (SEIFA) 2021*. Accessed: "20 July 2024". `https://www.abs.gov.au/websitedbs/censushome.nsf/home/seifa`.

Australian Bureau of Statistics (2024). *TableBuilder: The Online Tool for Creating Your Own Tables Using Census Data*. Accessed: 2024-10-28. `https://www.abs.gov.au/websitedbs/censushome.nsf/home/tablebuilder`.

Victorian Tertiary Admissions Centre (n.d.). *How to apply for university in Victoria*. https://vtac.edu.au/about. https://vtac.edu.au/about. (Visited on 10/28/2024).