Epidemiological Modelling of Spatial Dependency of Mental and Legal Challenges in Far North Queensland

School of Electrical Engineering, Computing, and Mathematical Sciences
Curtin University
This report is submitted for Master of Computing unit COMP 6016
May 2025

DECLARATION

Statement of originality

This document is the result of my own work and includes nothing, which is the outcome of work done in collaboration except where specifically indicated in the text. It has not been previously submitted, in part or whole, to any university of institution for any degree, diploma, or other qualification.

URL to video: video

Word count

This thesis comprises 12379 words including references (12,000 – 15,000 limit)

Signed: Nidup Dorj

Date: 27 May 2025

Your name and full qualifications

Nidup Dorji

Master of Computing

Abstract

The increasing prevalence of mental health issues presents significant challenges to first responders and healthcare systems across Australia. In response, the Australian Federal Government allocated 13.2 billion to mental health services in the 2021–2022 financial year, reflecting a consistent upward trend in funding. In Queensland, data from the Emergency Examination Authority (EEA) provide valuable insights into the spatial patterns of mental health-related emergency department presentations. This study proposes a comprehensive framework for data integration and statistical modelling to investigate the spatial associations between EEA occurrences and potential confounding factors, including socio-economic, retail, and environmental variables.

Global Moran's I was used to assess the presence of spatial autocorrelation in EEA counts. The analysis revealed a significant positive spatial autocorrelation, with Moran's I value of $I = 0.1691$ (p-value = 0.002), which exceeds the expected value under the null hypothesis ($E[I] = -0.027$). A two-stage statistical modelling approach was implemented. First, Lasso regression was applied to identify relevant covariates by penalising and reducing non-informative coefficients. Second, a Conditional Autoregressive (CAR) model was fitted using the EEA count as the response variable and the covariates selected in the previous step.

At the postcode level, the results demonstrated that EEAs were significantly and positively associated with the proportion of the Aboriginal population, with a posterior mean of 0.428 and a 95% credible interval of (0.089, 0.601). Similarly, a positive association was observed with the number of liquor outlets, evidenced by a posterior mean of 0.187 and a 95% credible interval of (0.002, 0.372). On the contrary, EEAs were significantly and negatively associated with average income, which shows a posterior mean of -1.066 and a 95% credible interval of (-2.002, -0.110), and with the proportion of individuals employed in blue-collar occupations, with a posterior mean of -0.591 and a 95% credible interval of (-0.956, -0.226).

The findings of the research underscore the complexity of mental health emergencies and spatially distributed socio-demographic and environmental factors. Such insights can inform policy makers to plan strategic interventions method and optimise the resource allocations to regions or individuals who are in need.

Contents

LIST OF ABBREVIATIONS AND ACRONYMS

EEA – Emergency Examination Authority

ED – Emergency Department

QPS – Queensland Police Service

QAS – Queensland Ambulance Service

PHA – Public Health Act (2005)

HHS – Hospital and Health Services

ABS – Australian Bureau of Statistics

OLGR – Office of Liquor and Gambling Regulation

AHPRA – Australian Health Practitioner Regulation Agency

PBOAS – Pharmacy Business Ownership Administration System

GP – General Practitioner

GLM – Generalised Linear Model

LASSO – Least Absolute Shrinkage and Selection Operator

CAR – Conditional Autoregressive (model)

MCMC – Markov Chain Monte Carlo

INLA – Integrated Nested Laplace Approximation

CI – Credible Interval

SD – Standard Deviation

MLL – Marginal Log-Likelihood

RACGP – Royal Australian College of General Practitioners

1 Introduction

1.1 Background

Mental health issues have become a significant concern and have become increasingly prevalent in our society in recent years, putting enormous pressure on the economy and health infrastructures. According to Australian Institute of Health and Welfare, the government spent more than $13.2 billion in the year 2022-2023 for mental health related services, corresponding to an approximate 11% increase from the year 2018-2012 of $11.8 billion[1]. And in the year 2023-2024, mental health and related illnesses cost the government an estimated $691 million[1]. The increase in expenditure underscores the growing recognition of mental health as a critical public health concern in Australia. In addition, there has been growing pressure on health professionals and facilities to address the growing mental health-related issues as evidenced by an increase in the number of mental health presentations in emergency departments[2]. There were 2,87,419 admissions to the ED for mental health and related illness in the year in 2022-2023 and of which 52% were classified as *urgent* and 20% as *emergency* on the triage category.

The current infrastructure has become inadequate to effectively accommodate the diverse range of cases and is failing to provide care to patients on time at EDs. The federal government and the state governments have set a 4-h target for complete care in EDs[2], but yet most of the mental health patients are likely to wait longer, some waiting up 2 hours[1] and staying longer than 18 hours[1][2]. The situation is worse in the regional and rural parts of Australia, where accessibility to mental health care is limited. The higher number of suicides in rural areas compared to metropolitan regions indicates the greater mental burden faced by individuals in regional areas due to limited access to health care [3]. Furthermore, rural residents are less likely to seek mental health care due to social stigma, fear of negative judgement[3], and limited understanding about mental health, further exacerbating the problem.

An emergency department presentation occurs when a patient arrives in the emergency department, registers, and undergoes an assessment by healthcare professional to determine the urgency of their care needs, known as triage. In some cases, individuals are presented under an involuntary assessment order [4]. The terminology and application of such orders differ between states and countries [4], as each jurisdiction applies different legal frameworks and procedures. In Queensland, it is known as “Emergency Examination Authority“, and it is a legal mechanism [4] under chapter 4A of *The Public Health Act 2005*, that permits the Queensland Police Service (QPS) and Queensland Ambulance Service (QAS) officers to detain and take a person to a health facility usually a hospital ED against their will if the officers believed the person to be at immediate risk of serious harm and that their (the person’s) behaviour warrants “immediate health examination”[5]. It also describes the specific responsibilities of hospital staff, including detention of the patient to allow time for assessment. The purpose of EEA is to address any urgent situations such as suicides, intoxication, aggressive behaviours irrespective of mental illness or some other reasons. The EEA procedure is unique to Queensland, where the mental health function is governed by public health legislation, rather than mental health legislation as in the case of other Australian jurisdictions[6].

The empowerment of officers under PHA to issue EEA, authorise detention and transportation of individuals against their will to emergency departments is a serious intervention[7][6] and thus legal safeguards are built in the process. Officers must complete a detailed and standardised EEA form prior to transferring the individual to the ED. The data collected through EEA forms captures detailed demographic characteristics, such as age and gender, as well as the specific circumstances and behavioural indicators that led to the emergency intervention, including evidence of self-harm, aggression, and substance use. The data is maintained and stored by the respective hospitals where EEA occurs, and currently there is no centralised database management to use the data for insightful analysis. Without a comprehensive and centralised reposity of EEA data, it becomes far more challenging to understand the real-world impact mental health challenges on hospitals, first aid responders, and the community at large.

The EEA records contain valuable information and can be used for effective monitoring, evaluation, and improvement of existing mental systems in the EDs. The EEA records collected over time could be used to find patterns and changes in mental health presentations[8][6][9]. This would help identify regions deemed to have higher EEA and individuals to be at immediate risk during an episode involving drugs, alcohol and psychiatric symptoms. Furthermore, this would help the first responders such as police and ambulance services to understand behaviour cues for appropriate assessment. By analysing these behaviour patterns, stakeholders can create and develop an effective strategies for prevention, enhance training for first responders and design a target region intervention for substance misuse and mental health crisis.

A legislative change from the Mental Health Act 2000 to the Public Health Act 2005 and the introduction of the EEA under that PHA, has reduced the public availability of data [9]. There has also been limited research focused on understanding and analysing EEA records. Much of the information contained in the EEA records remains unused [7], and even when it is used, it is mostly for qualitative analysis. While qualitative studies provide characteristics and different perspectives on mental health challenges, there is a lack of quantitative informational analysis and evidence to substantiate the findings and provide a comprehensive understanding of the patterns in data collected during EEA episodes. EEA records need to be consolidated with socioeconomic data from other sources to create a single repository and facilitate in-depth research on the effectiveness of crisis care, suicide prevention, and broader mental health system performance. This offers a unique opportunity to conduct quantitative analysis through statistical modelling to analyse and identify the spatial and socioeconomic confounding factors influencing the trends in the EEA episodes.

1.2 Problem Statement and significance

This study analysed and modelled the uncertainties of spatial dependency in endemic mental and legal challenges in far north Queensland by utilising medical information from the hospital and health services (HHS) during EEA episodes. We investigated what are the socioeconomic factors associated with EEA occurences and how does these factors, including household income and number of general practitioners, influence EEA admissions spatially. This to determine if the areas with higher liquor stores are associated with higher incidence of EEA episodes. A significant part of the study was also involved collecting, processing, and assimilating data from other sources with data from hospital data to create a mental health and epidemiology registry of database for far north Queensland.

The research was motivated by findings from the previous study “The Reasons You Believe”[7] and extended to incorporate large hospital dataset approximately 20,000 data points, along with data from bottle shops, general practitioners and pharmacies, to further explore the factors contributing EEA admissions to EDs. Through the data-drive modelling, the dataset and modelling would lead to effective and focused intervention strategies to mitigate circumstances surrounding the emergent mental health conditions. The master database system and modelling would pave the way to identify crisis ’hotpots’ and identification of spatial or demographic clusters as such age groups or specfic areas to have higher association of EEA admissions. This would have a huge implication on resource allocation and policy. Insights and trends analysis would help policymakers and health service planners make informed decision to allocate resources effectively to areas with higher need[7][9][6][8] and design specific interventions measures where crises are most frequent. Moreover, the findings will help evaluate the effectiveness of the current legal frameworks in Queensland—the only jurisdiction in Australia where mental health issues are governed under public health legislation—by measuring regional and temporal variations and identifying strengths and gaps that may warrant reform. With an effective data collection and analysis framework, this approach could facilitate a comprehensive understanding of mental health issues not only in Queensland but also allow for extension to other states, thereby informing targeted interventions and enhancing both immediate crisis care and long-term strategies for mental health system improvement.

1.3 Research Goals

1. Create a framework for data collection, data cleaning, data processing, and data assimilation in the context of mental health data with other data such as liquor data, pharmacies, to create a single incidence dataset.

2. Identify from aggregated data at the appropriate level (postcodes) that allows EEA monitoring and conduct exploratory analysis to detect potential association of report mental health and various confounding factors

3. Use machine learning methods and statistical modelling to understand statistical dependence, odds and risk of incidence for the various socio-economic factors at different postal regions.

2 Literature Review

2.1 Current State of the Art and Research Gaps

The Emergency Examination Authority records contain a wealth of information on mental health challenges in Far North Queensland. Despite their potential, the use of these data has remained and availability has limited[9][7]. With the growing burden of mental health problems manifesting in social, lifestyle, clinical, and economic domains, understanding spatial and geographical trends through statistical modelling and data-drive approaches has become increasingly paramount. Clear patterns of repeat admissions to emergency departments would allow healthcare professionals and first responders to develop more effective intervention strategies, alleviating the pressure of an increasing number of mental health-related ED presentations. The existence of spatial patterns in repeat EEA admissions was demonstrated by *Das et al.*[7] in their study "The Reasons You Believe...", which provided the foundation for this study.

The study collected 946 EEA forms from EDs from four Queensland Health and Hospital Services (HHS) districts from year 2017-2023 and demographic data (age, gender, population) from Australian Bureau of Statistics (ABS)[7]. An exploratory analysis and sentimental analysis of the merged data revealed variation in sentiment in by regions. The most commonly associated sentiments were fear and apprehension with words such as ”negative” and ”fear” appearing most in the EEA records while some regions also exhibited a higher association with positive sentiment with words as such ”positive” and ”trust” appearing most in the EEA records. A generalised linear (Poisson) regression model was deployed using predictable variables sentiment rage, age, sex ratio and outcome variable ”maximum count of repeat admissions”. A broader range of sentiment are found to be significantly associated with higher rates of repeated EEA admissions with unit increase in sentiments likely going to increase repeat EEAs by 1.34 times. In some remote and very remote northern parts of Queensland, the recurrence of EEA episodes is up to four times greater than in other regions[7]. A clear spatial clustering of repeat EEA admissions, indicating spatial differentiation in the nature and tone of mental health encounters. However, the sample was limited to non-metropolitan region in Queensland and no data from urban cities where recent research indicated many of the EEA presentations come from urban metropolitan areas[4]. Thus, constraining generalisability to the wider state population and to other states.

Alcock et al. [4] conducted a retrospective cohort study to examine the difference between adults brought to EDs by police with and without Emergency Examination Authority. The presentations brought in by police with EEA had a higher proportion of people with mental and behavioural disorders (52%) and were in the urgent category on the Australian Triage Scale [4]. They were also more likely to experience longer stays in the ED [2][4]. It also presented distinct demographic profiles. Presentation to EDs with EEA were likely to be female, less likely to be Aboriginal and/or Torres Strait Islander, more likely to live in major cities, and more commonly from socio-economically advantaged areas compared to those brought in without an EEA. Although most of the EEA presentations were from major cities, 15% of presentations were from regional, remote, or very remote areas, highlighting the burgeoning EEA presentations throughout the state. Presentations with missing EEA status were completely excluded, and analyses at the presentation level not individual level. Recurring admissions by the same person were not assessed during the study.

In a similar study, Clough et al.[6] identified significant temporal variations in the frequency and nature of EEA presentations, with peaks occurring during weekend nights, particularly between 11 p.m. and midnight. These findings align with broader patterns of high-alcohol-related incidents, where approximately 40% of assaults occur between 8:00 p.m. and 6:00 a.m. from Friday to Sunday[10]. Hospitals in Queensland experience the highest volume of ED presentations from evening through early morning on weekends and into Monday, a trend consistent across both major urban centres and smaller regional hospitals[6, 10]. Notably, many EEA presentations involve recurring admissions, with 23% of individuals having experienced two or more EEAs during the study period. Substance use emerged as the most common contributing factor, with 53% of cases involving substance-related issues. Moreover, the study found that 78% of individuals were discharged following examination, while only 22% were admitted or referred for further assessment. These findings highlight the need to better understand the factors influencing discharge decisions and the criteria for further assessment, which could improve evaluation and management processes in emergency settings. The study was limited to 10% of the EEA in the region and nearly half of the samples did not have complete information on the clinical outcome of the examination.

2.2 Summary and Research Position

All the existing studies on EEA records to date consistently highlight temporal, spatial, and demographic patterns in the occurrence of EEA episodes. However, many of these research works have analysed fewer than 1,000 EEA records, typically spanning only two to five years. For instance, Das et al. utilised 946 EEA forms from 2017 to 2022 [7], whereas Alcock et al. reported approximately 22,981 EEA presentations between 2018 and 2020—within a similar time frame. Likewise, Clough et al. [6] retrieved and analysed only 942 EEA forms, representing just 10% of all EEA presentations in the region during the observation period. As a result, the findings from existing studies may not be fully generalisable to other regions, as a substantial proportion of EEA records remain unexamined. This also highlights discrepancies in the datasets used across studies, largely due to limited access to comprehensive, statewide EEA data—further underscoring the need for a consolidated dataset to more accurately assess the spatial variation of mental health episodes across Queensland. Furthermore, none of the previous studies have explored the potential association between reported EEA presentations and various socio-economic factors. Spatial correlations involving confounding variables—such as the geographic distribution of bottle shops, general practitioners, and pharmacies—may have significant implications for understanding patterns in EEA occurrences. A more comprehensive study that includes additional covarying factors from a socioeconomic context is needed to map the spatial dependence of EEA episodes and address this gap in the literature.

This current study incorporated approximately 20,000 EEA records from 2008 to 2021, along with a range of socio-economic variables, including the spatial distribution of liquor stores, pharmacies, and general practitioners. Data from various sources were merged and assimilated into a unified repository. The initial phase of the exploratory analysis involved assessing spatial autocorrelation. Global spatial autocorrelation was evaluated using Moran's I statistic, while local spatial patterns were identified using Local Indicators of Spatial Association (LISA). The results indicated a significant positive spatial autocorrelation in EEA presentations. To investigate potential associations between EEA presentations and various confounding factors, a two-stage analysis was implemented. Firstly, Lasso regression was employed to reduce model complexity and prevent overfiting by selecting a subset of relevant predictors from the raw dataset. Secondly, a Conditional Autoregressive (CAR) Bayesian model was implemented to estimate spatial effects on EEA presentations. The CAR modelling approach facilitated detailed mapping and quantification of spatial variations and their interactions, while also accounting for spatial dependencies and reducing estimation bias.

Substantial studies, particularly in the field of epidemiology, have employed Conditional Autoregressive (CAR) models to account for spatial uncertainty and to estimate the odds ratio and relative risk of epidemics. Spatial and temporal models provide insights into the spatial and temporal variations in disease risk across a study area [11, 12]. Epidemics are typically spatially correlated; the disease risk in one area is often similar to that of neighbouring areas due to shared environmental, demographic, or socio-economic factors. Such spatial correlation is highly likely in the occurrence of EEA events in our study. Amsalu et al. [11] evaluated various spatial models to understand the spatio-temporal distribution of tuberculosis in the geriatric population of China. Their findings showed that the CAR model, which incorporated spatial and temporal structures along with their interactions, performed best in modelling the spatio-temporal data.

We deployed a Conditional Autoregressive (CAR) model to examine the spatial patterns of EEA occurrences in Far North Queensland and to evaluate their relationship with social and economic factors. This is a comprehensive analysis that accounts for a broader range of variables than previous studies and incorporates a large dataset of EEA records. The findings from this study will complement and strengthen the results of [7], providing valuable insights for strategic resource allocation in regions with a higher recurrence of EEA episodes. Such targeted interventions have the potential to alleviate pressure on emergency departments (EDs) and the first responders.

3 Methods

3.1 Setting

The study was conducted in the Far Northern region of Queensland, encompassing remote and non-metropolitan areas of the state. This region represents approximately 15% of Queensland's total population of 5.2 million people [13] (see Appendix 8.2). Compared to metropolitan areas, these regional, remote, and very remote communities have limited access to mental health services. Residents are also less likely to seek mental health support, due to factors such as social stigma [3], negative community sentiments [7], and a general lack of awareness regarding available health resources. This setting reflects the real-world complexities and challenges of addressing mental health issues in areas where service accessibility is severely constrained by geographic and social barriers.

3.2 Sampling

A total of 22,429 EEA records were collected from the Hospital and Health Services (HHS) of four Queensland districts: Mackay, Townsville, Far North, and Mount Isa. The 20000+ EEA data account for 86 postal regions from these four districtions out of 427 total postcodes in Queensland. Two postcodes were excluded from the analysis: postcode 4803, due to the absence of corresponding shapefile and postcode 4825, as it extends across state or territory boundaries. The study sample included 84 postcodes from the four districts. In addition to EEA data, the study incorporated contextual information, including 9,731 liquor store records, 1,237 schools, 1,309 pharmacies, 337 libraries, and 1,391 general practitioners. Socio-economic factors for the 84 postcodes were obtained from the Australian Bureau of Statistics (ABS).

3.3 Ethics and Data Privacy

Ethical approvals were obtained from the relevant ethics committees, including the Townsville Hospital and Health Services Human Research Ethics Committee (HREC Reference number: LNR/2018/QTHS/46061), and James Cook University (HREC Reference number: H7672).

Data used in this research, except the EEA records, are obtained from publicly available stores. The study used only administrative data without any human participants; therefore, no informed consent was required for data collection. Additionally, no personally identifiable data was accessed, used, or stored at any state during the research.

State-of-the-art protocols were followed for data management systems considering technological aspects, data security, privacy and other ethical issues following the Queensland Health Standard[14] and Data Management policy[15], which recognise data as the strategic possession of the Department of Health and outlines the requirements of consistent and effective data management to support health services and outcomes.

3.4 Data Collection

A comprehensive data collection method was implemented to collect medical, administrative , retail, and socioeconomic data from various sources. While most of the data can be found publicly available, some of the them are restricted for use and require permission to use. A few of the retail data were collected through phone calls and emails, while others are scrapped from the internet.

Figure 1 shows the comprehensive flow of data processing, data aggregation, and statistical modelling used in the study.

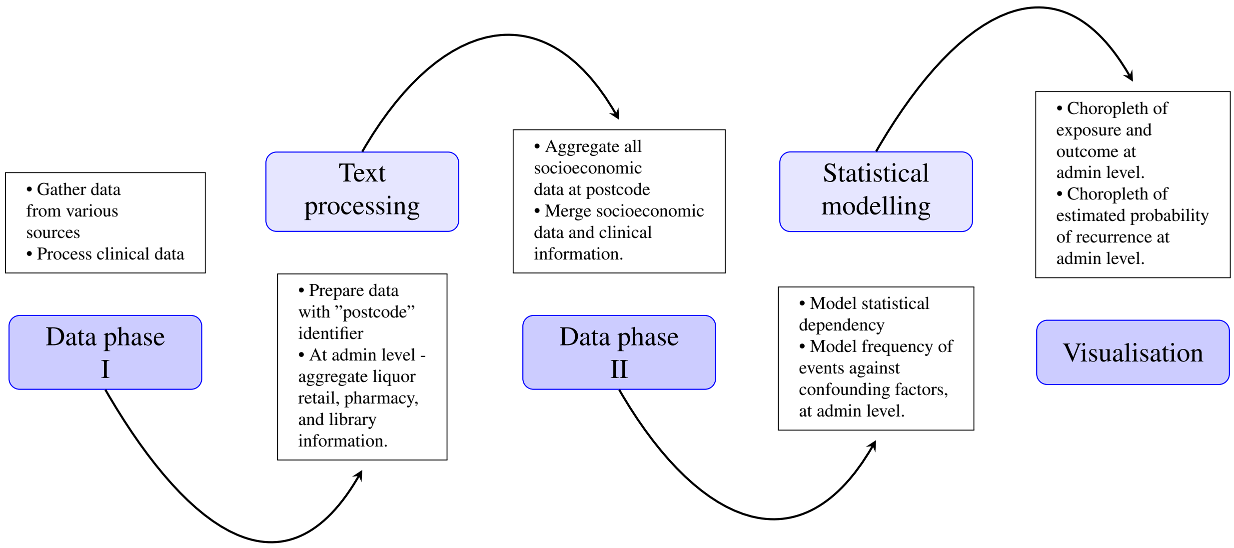


Figure 1: Flowchart of data collection, aggregation and statistical analysis.

3.4.1 Data Sources

The data for this were obtained from two different types of sources: (1) agencies and (2) online sources. A significant portion of data is regulated by various state different government departments and other regulatory bodies. They collect routinely and maintain extensive dataset. While most of the data can be viewed on their public websites, it is often not readily available for direct download. Access typically requires formal permission, and in some instances, involves associated costs. Various inquiries were made to determine the availability of data and retrieval options to the relevant agencies. The following agencies were contacted for data collection:

- Hospitals and Health Services
- Office of Liquor and Gambling Regulation(OLGR)[16]
- Queensland Health
- Australian Health Practitioner Regulation Agency[17](AHPRA)
- Queensland Government

The study relied on the Australian Bureau of Statistics (ABS) for socioeconomic and census data. The ABS online portal provided a user-friendly interface that facilitated the retrieval and download of various datasets. However, the ABS does not host retail data. Some of the retail data were directly accessible and available from the Internet without the need to contact the agencies. The following online sources were used for data collection

- Australian Bureau of Statistics (ABS)
- Queensland Government Open Data Portal [18]

- Queensland Government Pharmacy Business Ownership Administration System (PBOAS) Portal
- State Library of Queensland [19]
- Australian Commission on Safety and Quality in Healthcare[20]

The EEA forms obtained from the Hospital and Health Services (HHS) included medical, clinical, and demographic information related to the patients. The Office of Liquor and Gaming Regulation (OLGR) provided liquor licensing data, which included store addresses, license types, and operating hours. This dataset was comprehensive, encompassing all categories of liquor outlets, including both on-premise and off-premise establishments. Data on general practitioners were scraped from the Australian Commission on Safety and Quality in Health Care website. Queensland Health provided pharmacy data. Information on schools and libraries was retrieved from the Australian Government Open Data Portal. Socio-economic data, including personal income and employment status, were sourced from the Australian Bureau of Statistics (ABS).

3.4.2 Data Collection Methods

To collect data that were not publicly available for direct download, two data acquisition methods were used:

- **Phone and Email Outreach**: Most government departments and regulatory bodies provide contact information on their official online portals. Initial phone calls were made to inquire about the availability of relevant data, followed by formal email requests using a standardised template (see Appendix 8.2). We specifically requested the data in CSV format to ensure compatibility with other datasets used in the study.
- **Web Scraping from Online Sources**: In certain cases, data were available on official websites but could not be downloaded directly. For example, the Australian Commission on Safety and Quality in Health Care hosts detailed information on general practitioners across Australia. To extract this data, a web scraping method was implemented using a Python script and Selenium automation framework [21] (see Appendix 8.3).

3.5 Data Preprocessing

3.5.1 School and Library Data

The Queensland Government Open Data Portal contained massive repository of over 175,000 datasets. A total of 1,774 school and 337 library data were obtained from it. However, for the study focused on 84 postcodes in far north Queensland that covered only four districts, a subset of 266 libraries and 100 libraries was included.

3.5.2 Pharmacy Data

A total of 1309 pharmacy stores data acquired from the Queensland Health website. Other agencies like the the Australian Health Practitioner Regulation Agency (AHPRA) provides an option to view information about pharmacists and pharmacy locations across different postal regions through its website. However, that data is not accessible to download of their website directly and it was informed that access to the data required payment. Instead, the most recent PDF report[22] of licensed pharmacies was obtained from Queensland Health. This report contains pharmacy data sourced from the Department's Pharmacy Business Ownership Administrative System (PBOAS), which is governed by the Pharmacy Business Ownership Act 2001 (PBO Act). The PDF was converted into a spreadsheet using Microsoft Excel [23], and subsequently transformed into CSV format. Of the 1,309 pharmacy records collected, only 196 corresponded to the 84 postcodes included for analysis.

3.5.3 Liquor Data

The *Office of Liquor and Gambling Regulation* provided a comprehensive dataset containing 9,986 liquor license records across Queensland, encompassing a wide range of establishments—from commercial wholesalers to artisan producers. OLGR regulates liquor and wine licences and permits according to the Liquor Act 1992. The dataset contained numerous premises where the sale of liquor was a secondary function, such as restaurants and motels. These were not included in the data as the focus was solely on the establishments whose primary function involved direct sale to consumers. Based on OLGR's licence category descriptions, only the following license types were retained:

1. Commercial Other - Producer/Wholesaler
2. Commercial Other - Industrial Canteen
3. Wine Producer
4. Wine Merchant
5. Commerical Other - Bar
6. Nightclub
7. Commerical Other - Artisan Producer

3.5.4 General Practice

General Practitioners data for whole of Australia were available on the Australian Commission on Safety and Quality in Health Care website. The website hold 6610 records of accredited general practices and aboriginal medical services by the Royal Australian College of General Practitioners (RACGP) standards. However, it was not accessible to download; therefore, the data was extracted from the website using Python and Selenium automation framework(refer to Appendix 8.3). A total of 1391 records of general practitioners and aboriginal medical services based in Queensland were extracted and processed. Figure 2 shows structure of the extracted data.

	name	category	address	suburb	state	postcode	date
0	121 Medical Centre	General practice	6 / 121 Shute Harbour Road	CANNONVALE	Qld	4802	06-Jul-2026
1	188 Medical	General practice	188 Brisbane Road	ARUNDEL	Qld	4214	08-Jan-2028
2	19th Avenue Family Practice	General practice	Shop 17, 19th Avenue Shopping Centre	Elanora	Qld	4221	09-Aug-2025
3	7 Springs Medical Practice	General practice	881 Ruthven Street	Toowoomba	Qld	4350	19-Oct-2026
4	Abbott Medical Clinic	General practice	3 / 473 Mulgrave Road	EARLVILLE	Qld	4870	27-Jul-2025

Figure 2: General Practitioner Data extracted from Australian Commission on Safety and Quality in Health Care

3.5.5 Data from ABS

The Australian Bureau of Statistics (ABS), a national agency, publishes many datasets, including economic, social, population, and environmental data for research and policy making at different administrative levels. Although datasets in ABS are raw and not formatted, the TableBuilder tool allows one to create table data from different categories into customised format for specific purposes.

For this study, socioeconomic data was extracted from the 2019 Census of Population and Housing Data, particularly focusing on variables related to employment, income, and education. The census data include counts for a variety of demographic and socioeconomic indicators, such as indigenous status, age (in 5-year intervals), and weekly income ranges. The following variables were selected for analysis, as defined in the ABS Census Dictionary (2021) [13]

1. Indigenous Status (INGP): categorised into non-Indigenous, Aboriginal, Torres Strait Islander, both Aboriginal and Torres Strait Islander, or not stated.
2. Age Group (AGEP): categorised into different age intervals.
3. Highest Year of School Completed (HSCP) data: categorised into completed year 12, 11, 10, 9, 8, 'did not go to school'
4. Occupation: various occupational classifications as defined in ABS.
5. Total Personal Income (weekly): categorised into defined weekly income intervals.
6. Population data: total population per postal region.

All of these variables were grouped at the postal level (postcode) in TableBuilder and downloaded separately for the final aggregation with EEA records and retail data.

Each dataset from ABS contained many categories or intervals, resulting in a high-dimensional feature space with over 90 features during the final aggregation. This complication required using feature engineering method to reduce the number of dimensionality and create concise and interpretable categories. The age group data was summarised into three categories aggregated at postcode level: $<$ 18 years; 18-64; and age $>=$ 65 years similar to Das et al[7]. Likewise, the highest year of school completed data was also summarised into three categories: $<$ year 8; year 9-12 and school not attended.

The occupation classifications were grouped into blue vs white collar occupation. Hu et al.[24] defined blue-collar workers as those who perform primarily physical work and those career paths are relatively restricted while white-collar workers are professional and semi-professional employees whose nature of work is primarily mental or administrative work. Based on this, technicians and trades workers, community and personal service workers, machinery operators and drivers, and labourers were categorised into blue-collar occupation. White-collar occupation includes managers, professionals, clerical and administrative workers, and sales worker.

The weekly income variable comprises of 14 income range brackets. Instead of using all different range, the average weekly income was calculated by taking the median of the income range bracket, multiplying it by the corresponding frequency and dividing by adjusted population, as shown in Equation 1. For example, the range "$1 - $149" is assigned a midpoint of $75 and this is to ensure uniform distribution within each range. The adjusted population excludes individuals for whom income was not stated, as well as those reporting nil or negative income. Table 1 shows the new columns in the weekly income variable extracted from ABS.

$$\text{Average Income} = \frac{\sum(\text{frequency} \times \text{midpoint})}{\text{adjusted total population}} \tag{1}$$

postcode	negative_income	nil_income	average_income
4000	76	1733	1344.000000
4005	65	676	1691.000000
4006	58	1017	1431.000000
4007	49	970	1605.000000
4008	0	12	1175.000000

Table 1: Weekly Income data extracted from ABS and recategorised

3.5.6 EEA Data

EEA data contained 22429 EEA records generated by the Queensland Police Service or the Queensland Ambulance Service and handed over to the Emergency Departments in four HHS districts from 2008-2020. Each record contained 42 attributes, including a unique occurrence ID, occurrence address, timestamp, drug and alcohol involvement, and year of occurrence.
As this study focused on the frequency of EEA occurrences at the postcode level and their association with potential confounding factors, only two variables—occurrence ID and postcode—were retained for analysis. All other variables were excluded and no temporal variation was considered. The data were then aggregated by postcode to generate the total count of EEA occurrences per postcode.

	postcode	eea_count
0	4482	2
1	4701	1
2	4707	7
3	4721	32
4	4723	1
5	4737	171
6	4738	9
7	4739	6
8	4740	1995
9	4741	52

Table 2: First 10 rows of EEA data after preprocessing

3.6 Data Aggregation

3.6.1 ABS, Retail and EEA data Merge

All the datasets including retail data, ABS data, and EEA records were aggregated at postcode level. School, pharmacy, liquor stores, libraries and general practitioner data were converted to frequency data, with counts aggregated by postcode. The ABS datasets were already aggregated at the postcode level. An outer merge technique is used to aggregate all the data sources at the postcode. The methods retains even the missing values across variables, which are calculated in the subsequent data processing method. Table 3 shows aggregation of all the retail datasets at postcode level.

postcode	school	liquor	average hours	pharmacy	library	gp
2029	NaN	1.000000	15.000000	NaN	NaN	NaN
2100	NaN	1.000000	14.000000	NaN	NaN	NaN
2406	NaN	1.000000	14.000000	NaN	NaN	NaN
2880	NaN	1.000000	14.000000	NaN	NaN	NaN
2899	NaN	NaN	NaN	NaN	NaN	1.000000
4000	1.000000	332.000000	14.289157	24.000000	2.000000	16.000000
4001	NaN	4.000000	14.000000	NaN	1.000000	1.000000
4004	NaN	11.000000	14.000000	1.000000	NaN	NaN
4005	1.000000	60.000000	13.533333	3.000000	1.000000	2.000000
4006	NaN	317.000000	14.444795	13.000000	NaN	9.000000

Table 3: Retail data aggregated at postcode

3.6.2 Data Merge Validation

The data used in this study were obtained from multiple sources, rather than a single unified dataset. Several of these datasets required extensive pre-processing, including the removal of incomplete or irrelevant information. There was considerable likelihood of inaccuracies or errors in the datasets due to heterogeneity of the sources and complexity of the preprocessing steps. This would subsequently result in an erroneous data assimilation.

To mitigate this risk and verify the accuracy of the aggregated dataset, an automated random merge validation procedure was implemented in Python (see Appendix 8.4). The validation function accepts both the final merged dataset and an individual unmerged dataset- such as one of the ABS datasets or retail datasets as input. Then it randomly selects five rows from the unmerged dataset and checks for consistency by comparing them to their corresponding entries in the merged file. This ensured the consistency and correctness of the aggregated data.

3.7 Data Preparation

3.7.1 Statistical Imputation

The aggregated data contained 84 postcodes and 24 variables in a single file. However, due to the use of an outer aggregation method, the dataset contained multiple missing (NA) values. These missing values could not be removed, as they affected more than one-third of the 84 postal regions. Specifically for pharmacies, libraries, and general practitioners, which had 21, 25, and 26 null values, respectively.

Although these missing values could have been populated using the average for each variable, this approach risked introducing bias and producing non-representative estimates, as many postcodes would be assigned identical values. To address this issue more rigorously, a negative binomial generalized linear model (GLM) was used to replace the missing values. The negative binomial regression was appropriate given the observed overdispersion, where variance

	NegBin GLM
(Intercept)	-5.12^{***}
	(0.75)
log(population)	0.69^{***}
	(0.08)
AIC	224.42
BIC	230.60
Log Likelihood	-109.21
Deviance	40.24
Num. obs.	58

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 4: Negative Binomial GLM for predicting GP null values

approximately equal the squared mean. Since the availability of pharmacies, libraries, and general practitioners is often associated with population, the generalised linear model was fitted with non-null observations as response variable and population as predictor variable. The fitted model subsequently generated predictions for missing values based on postcode-level population data. This approach yielded statistically informed estimates while maintaining spatial heterogeneity in the replaced values.
Table 4 shows the negative binomial generalised linear model used to populate the null-values of general practitioners based on the population at postcode level.

3.7.2 Variables for modelling

The finalised dataset contained 22 variables with no missing values following pre-processing, data cleaning, and imputation methods. These variables encompass a comprehensive range of socio-demographic, economic, and community characteristics, providing a multifaceted perspective on Emergency Examination Authority events. The following outlines the detailed operational definitions of each variable.

1. **eea_count**: Total number of Emergency Examination Authority (EEA) cases recorded within each postcode between 2008 and 2020; this serves as the response variable.

2. **non_indigenous**: Number of individuals who do not identify as Aboriginal or Torres Strait Islander in each postcode.

3. **aboriginal**: Number of individuals identifying as Aboriginal.

4. **torres_strait**: Number of individuals identifying as Torres Strait Islander.

5. **both_aboriginal_and_torres_strait**: Number of individuals identifying as both Aboriginal and Torres Strait Islander.

6. **age_18_below**: Number of individuals aged 18 years and younger.

7. **age_18_64**: Number of individuals aged between 18 and 64 years.

8. **age_65_above**: Number of individuals aged 65 years and above, representing the elderly population.

9. **school_years_9_12_attended**: Number of individuals who have completed secondary education between Years 9 and 12.
10. **year_8_below**: Number of individuals who completed education up to Year 8 or below.
11. **school_not_attended**: Number of individuals who did not attend school.
12. **negative_income**: Number of individuals reporting a financial loss (negative income) in the 2021 Census.
13. **nil_income**: Number of individuals reporting no income.
14. **average_income**: Average weekly income in Australian dollars, estimated using midpoint values of income brackets and adjusted for population.
15. **blue_collar_occupation**: Number of individuals employed in manual or trade-related occupations.
16. **white_collar_occupation**: Number of individuals employed in professional, managerial, or administrative roles.
17. **population**: Total resident population within each postcode, used as an offset variable in statistical modelling.
18. **school_count**: Number of schools located in each postcode.
19. **liquor_count**: Number of liquor stores in each postcode.
20. **pharmacy_count**: Number of pharmacies in each postcode.
21. **library_count**: Number of libraries in each postcode.
22. **gp_count**: Number of general practitioners (GPs) in each postcode.

3.8 Monotonic Log Transformation of Covariates

All the variables in this analysis are count variables and exhibit skewed distributions, as shown in Figures 3 and 4 of the few selected variables. Most of them are positively skewed typical in count data, with the majority of observations concentrated at lower values. Such distributions across predictor variables can compromise model performance, potentially leading to biased estimates and reduced accuracy.

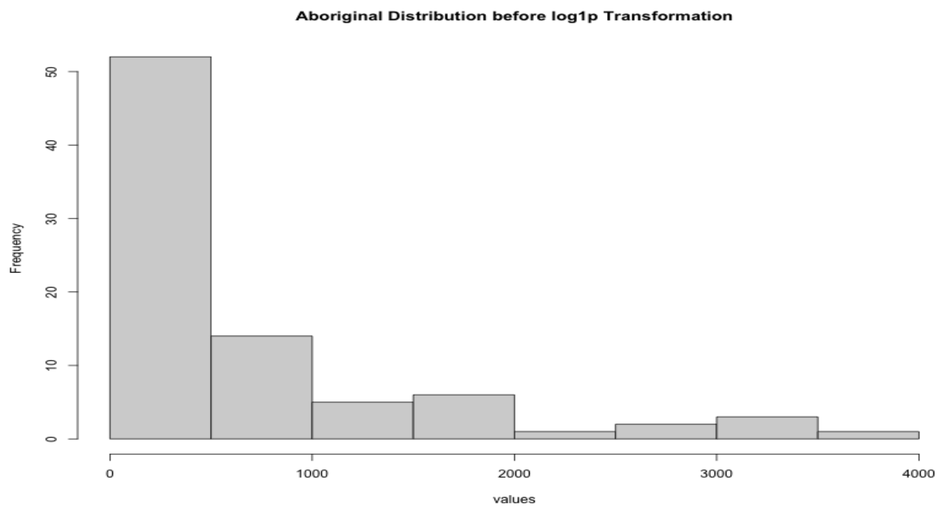


Figure 3: Aboriginal Data Distribution Before Log Transformation

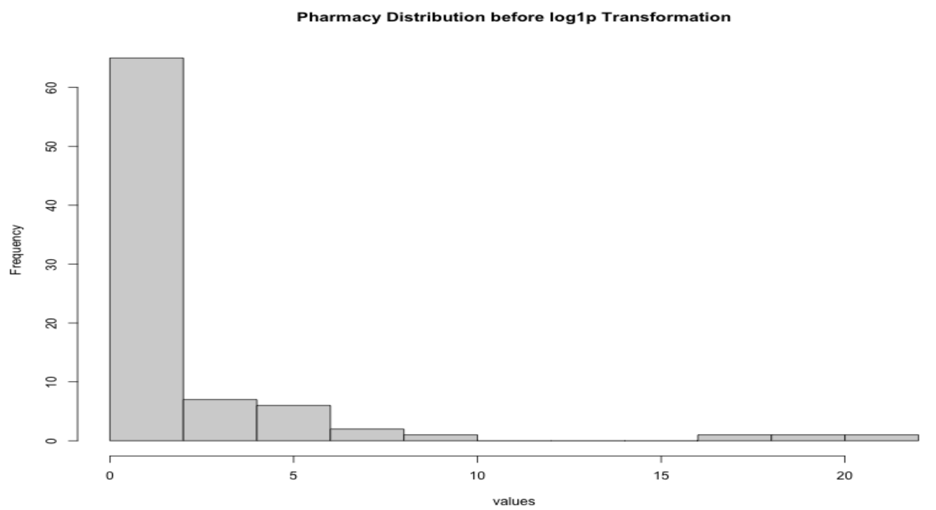


Figure 4: Pharmacy Data Distribution Before Log Transformation

To address this, all predictor variables were log transformed using `log1p` function in R[25], defined as $\ln(x+1)$. Figure 5 and 6 shows data distribution after applying `log1p` transformation. This transformation effectively reduces skewness and produces a more symmetric distribution. The `log1p` is particularly suitable for count values as it accommodates variables with 0 counts without undefined error (e.g., $\log(0)$). Furthermore, monotonic log transformation preserves the relationship between the response variable and the predictors by maintaining spatial variations and trends of relative risk.

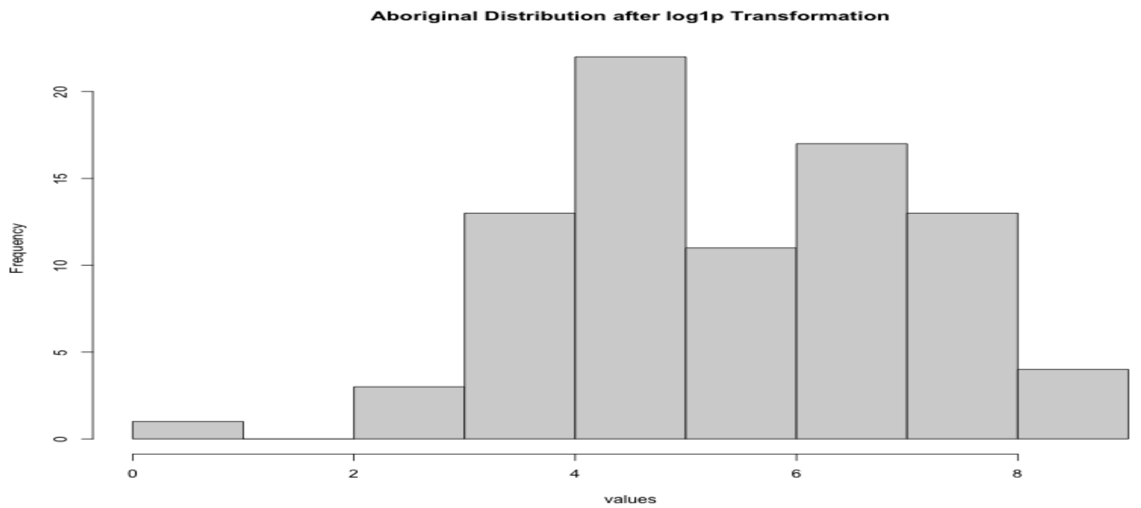


Figure 5: Aboriginal Data Distribution after Log Transformation

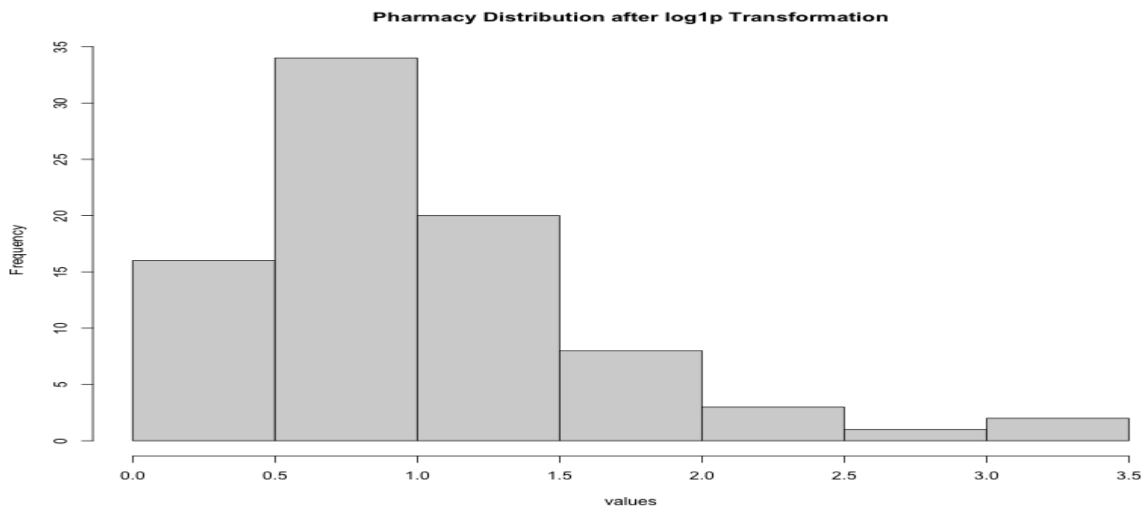


Figure 6: Pharmacy Data Distribution after Log Transformation

3.9 Statistical Analysis

A suite of R packages was used to conduct the spatial and statistical analyses. R offers a comprehensive ecosystem of packages that facilitate modelling without requiring in-depth understanding of the underlying computational complexities. For data processing, wrangling, and analysis, the **dplyr** package was utilised. Spatial data manipulation and analysis were performed using the **sf** package. Visualisations were generated using **ggplot2**, with advanced visualisations done using **plotly** and **mapview** for more advanced and interactive maps.
To test for the presence of spatial autocorrelation in EEA counts, the **spdep** package[26, 27, 28, 29] with Moran's I statistics was used. For baseline statistical models, generalised linear models were implemented using the **glmnet** package, which also supports Lasso regression. To model spatial dependencies, the **INLA** (Integrated Nested Laplace Approximation)[30, 31, 32, 33] package was used due to its computational efficiency and suitability for Bayesian spatial modelling.

3.9.1 Spatial Distribution of EEA

Moran's I was used to investigate the spatial autocorrelation of the EEAs. Moran's I describe the extent to which a variable is correlated with itself through the space. It quantifies how similar or dissimilar each postal region is with its neighbours, and averages the spatial correlation over the across the entire study area. The mathematical definition of the global Moran's I[34] is shown in Equation 2. The statistic is bounded between -1 and 1, where values near 1 indicate strong positive spatial autocorrelation, values near -1 indicate strong negative spatial autocorrelation, and values near 0 suggest spatial randomness.

$$I = \frac{N}{W} \cdot \frac{\sum_i \sum_j w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2} \tag{2}$$

Where:

N $\rightarrow$ Number of regions (84 postcodes)

x_i $\rightarrow$ Observed value of the variable of interest at region i (EEA count)

$\bar{x}$ $\rightarrow$ Mean of all values x

w_{ij} $\rightarrow$ Spatial weight or proximity between regions i and j (e.g., 1 if they are neighbours, 0 otherwise)

W $\sum_i \sum_j w_{ij}$ = Sum of all spatial weights such that $i \neq j$

The package **spdep** in R was used to calculate the Moran's I of EEA as follows. The postal area shape file was imported from ABS and merged with the analytical dataset. The neighbours for each postal region is calculated using *poly2nb()* of spdep package, which uses queens (8 neighbour rule) continuity to determine the neighbours. Then the list of neighbours is converted to a spatial weight matrix using *nb2listw()*. The resulting list of neighbours was then converted into a spatial weights matrix using *nb2listw()*. This spatial weights matrix was passed to the *moran.test()* function, along with a specified hypothesis, to calculate Moran's I statistic.

To test the presence of spatial autocorrelation, the computed Moran's I is then compared to its expected value under the null hypothesis (no spatial autocorrelation), which assumes that observations x_i are independent and identically distributed (i.i.d.) and that Moran's I follows a normal distribution [12]. The expected value of Moran's I under the null hypothesis is given by $E(I) = -1/(n-1)$, where n is the number of spatial units. Moran's I values greater than $E(I)$ suggest positive spatial autocorrelation, the values less than $E(I)$ indicate negative spatial autocorrelation, and values approximately equal to $E(I)$ imply spatial randomness.

3.9.2 Epidemiological Modelling

The modelling was implemented in two stages. Firstly, a (Lasso) Least Absolute Shrinkage and Selection Operator regression is implemented to identify and retain all the predictors that are likely to be associated with EEA occurrences. Lasso applies absolute shrinkage penalities to regression coefficients and performs feature selection by shrinking the coefficients of the irrelevant coefficients to zero. To get a optimal regularisation parameter that minimises the mean squared error, a ten-fold cross-validation method with maximum iteration of 10^6 was used. Given that the response variable—EEA count—is a count variable, it was assumed to follow a Poisson distribution. Additionally, population of each postcode was included as an offset term to account for varying population sizes across regions.

A hierarchical Conditional Autoregressive (CAR) model was used to model the response variable, EEA count, against the predictors selected through the Lasso regression. This modelling was implemented using the `INLA` package in `R`. The Integrated Nested Laplace Approximation (INLA) is a computationally efficient alternative to traditional Markov Chain Monte Carlo (MCMC) methods, offering substantial improvements in computational speed[35, 36].

CAR models are suitable for modelling spatial autocorrelation in small areal or lattice data[36] and evaluating the affect of spatially structured random effects on the response variable. Besag model with intrinsic conditional autoregressive structure was used to capture the affect of the spatially structure random effects based on the neighbourhood adjacency relationships. For our study, the CAR model estimates the likelihood that EEA counts in neighbouring postcodes are likely to be correlated due to shared characteristics, environmental factors, or unmeasured confounding factors that exhibit spatial clustering. This is achieved through Bayesian inference to estimate the posterior distribution of the covariates using INLA.

Upon analysis, the assumption of a Poisson distribution for EEA count appeared to be inappropriate, as the data exhibited overdispersion, where the variance was approximately equal to the squared mean. Thus, four different models were fitted as follows.

- **Model 1:** A baseline Generalised Linear Model (GLM) assuming a Poisson distribution, without accounting for spatial autocorrelation or spatial effects.
- **Model 2:** A baseline Negative Binomial GLM that does not consider spatial autocorrelation but accounts for overdispersion in the data.
- **Model 3:** A Conditional Autoregressive (CAR) model assuming a Poisson distribution, incorporating spatial autocorrelation and spatially structured random effects.
- **Model 4:** A Negative Binomial CAR model that incorporates both spatial autocorrelation and spatial effects, while also addressing overdispersion.

The following equation 3 explains the model structure with the assumption that EEA count follows a Poisson distribution with mean λ.

$$\begin{aligned} y_i &\sim \text{Poisson}(\lambda_i), \\ \log(\lambda_i) &= \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} + s_i, \quad i = 1, 2, \ldots, N \end{aligned} \tag{3}$$

where λ_i denotes the expected value at the i^{th} postal region; β_0 is the intercept; β_j are the coefficients of the covariates x_{ij} in region i for i=1, 2, ..., 84, j = 1, 2, ... 20 ; and s_i denotes the spatially correlated random effect for region i, which follows an conditional autoregressive distribution.

Similarly to the Lasso regression approach, the `poly2nb()` function from the `spdep` package was used to compute the neighbours of each postal region based on the Queensland shapefile. An adjacency list was then generated using the `nb2INLA()` function, which produces a graph file encoded with the neighbourhood graph structure compatible with the `INLA` package. This adjacency graph serves as the foundation for specifying the conditional autoregressive (CAR) prior for the spatially structured random effect.

The four models are then fitted to calculate the fixed effects (socio-demographic covariates selected from Lasso regression and population offset) and spatial random effects. Summary statistics were obtained for the posterior distributions of each parameter, including the posterior mean, standard deviation, and the 2.5^{th} and 97.5^{th} percentiles, which together define the 95% credible interval. The marginal log-likelihood dictated model selection. The model with the higher (i.e., less negative) marginal log-likelihood was the preferred model.

4 Results

A total of 22,429 EEA records across 84 postal regions in Queensland were analysed to examine spatial dependency in relation to a range of contextual variables, including socioeconomic, retail, and pharmaceutical factors. Prior to statistical modelling, global Moran's I was applied to assess the presence of spatial autocorrelation in EEA counts. Upon identifying significant spatial autocorrelation, a two-stage modelling approach was implemented. In the first stage, Lasso regression was implemented to subset covariates that are potentially associated with spatial variation in EEA counts. Then multiple statistical models were fitted including baseline GLMs and CAR models to evaluate spatial effects on EEA accounts.

4.1 Exploratory Analysis and Pattern of clustering

Fig 7 shows the spatial distribution of the EEA and a selection of covariates. The exploratory analysis was conducted after applying a `log1p` transformation to address skewness in the original data; however, the monotonic nature of the transformation preserved the underlying trends, as confirmed by comparisons before and after transformation. As expected, the analysis reveals spatial autocorrelation in both EEA counts and the covariates, with postal regions of similar values clustered together.

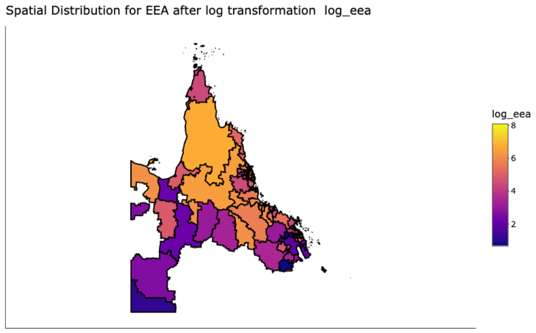
Notably, the analysis provides evidence of an association between EEA occurrences and potential confounding factors. For instance, postcode 4870 (Cairns) recorded the highest number of EEA cases (3,266) during the study period. This region also had the highest number of liquor stores (341), the second highest number of individuals with nil income (3,736), a substantial Aboriginal population (3,422), and the second highest number of individuals engaged in blue-collar occupations (15,247). These patterns suggest a strong spatial association between EEA occurrences and underlying socio-economic, retail, and environmental factors.

4.2 Spatial Distribution of EEA counts

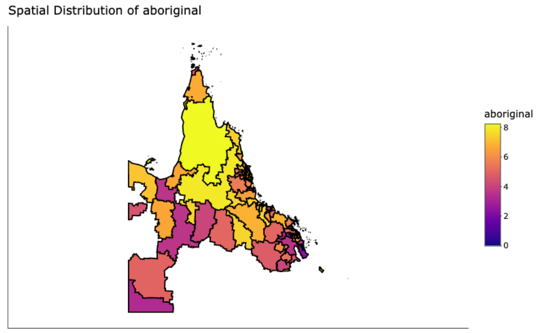
Table 5 shows the Moran's I statistic calculated for the observed EEA count data. The spatial autocorrelation analysis indicates that the distribution of EEA counts is spatially correlated across the 84 postal regions, with a global Moran's I value of $I = 0.1691$. This value is substantially greater than the expected value of I under the null hypothesis ($E[I] = -0.0127$), suggesting the presence of positive spatial autocorrelation. As a result, the null hypothesis of no spatial autocorrelation is rejected. Furthermore, the associated p-value of 0.002 confirms that the observed spatial clustering is statistically significant at the 5% level. These findings imply that the clustered pattern of EEA occurrences is unlikely to be due to random chance, indicating a degree of spatial dependency in the distribution of EEAs in far north Queensland.

Figure 8 demonstrates the Moran's I scatterplot for EEA, where each point represents a spatial unit (postcode). The scatterplot displays the EEA counts on the x-axis against their spatially lagged values on the y-axis, which are the weighted averages of EEA counts in neighbouring regions. The slope, Moran's I, indicates the strength of the and direction of the spatial autocorrelation. The clusterings of the spatial units near the first quadrant supports the presence of spatially clustered regions with similar EEA counts (high-high or low-low).

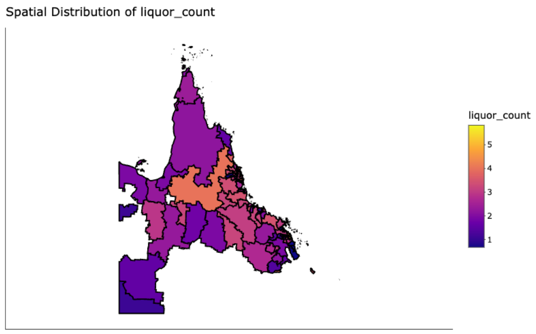
Given the substantial variation in population sizes and EEA counts across different postal regions, as a standard practice in epidemiological studies, a standardised EEA ratio was calculated



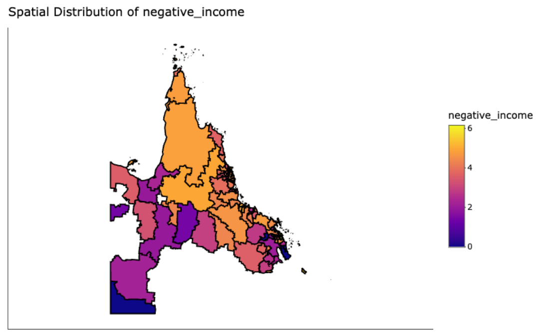
(a) EEA count



(b) Aboriginal Count



(c) Liquor frequency



(d) Negative income

Figure 7: Spatial variation of EEA counts and few selected covariates across the study region.

Test Description	**Value**
Data	`data$eea_count`
Weights	`nbw`
Adjustment	n reduced by no-neighbour observations
Moran's I Statistic	0.1691
Expectation under H_0	-0.0127
Variance	0.0045
Standard Deviate	2.7037
p-value	0.0034
Alternative Hypothesis	Greater (positive spatial autocorrelation)

Table 5: Global Moran's I Test under Randomisation for EEA

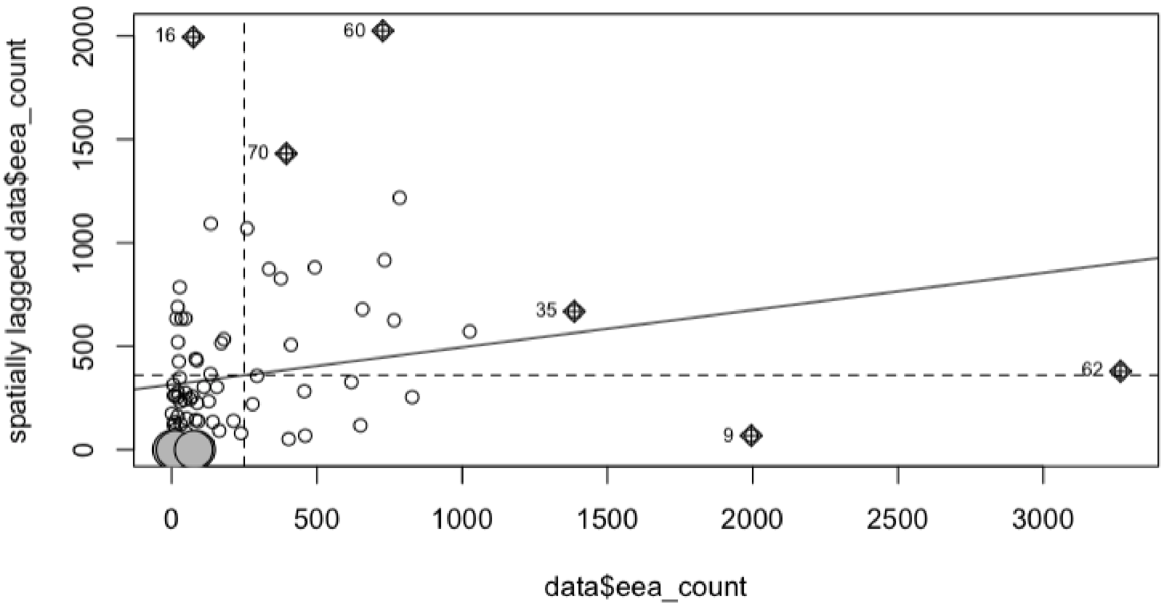


Figure 8: Moran's I Scatterplot

by expressing EEAs in person 10000 people in each postal region. This standardisation allows a clearer and more meaningful comparison of spatial effects on a common scale. Table 6 shows the Moran's I statistics for the standardised EEA ratio. Consistent with our previous finding, the result exhibited positive spatial autocorrelation with $I > E[i]$ and $p - value < 0.05$.

Test Description	**Value**
Data	`data$standardised_eea_ratio`
Weights	`nbw`
Adjustment	n reduced by no-neighbour observations
Moran's I Statistic	0.1085
Expectation under H_0	-0.0127
Variance	0.0037
Standard Deviate	1.9883
p-value	0.0234
Alternative Hypothesis	Greater (positive spatial autocorrelation)

Table 6: Global Moran's I Test under Randomisation for Standardise EEA ratio

4.3 Epidemiological Modelling

4.3.1 Subset selection using Lasso

Figure9 shows the number of covariates that lead the minimum mean-cross validated error during Lasso feature selection. Out of 20 covariates, Lasso selected 10 features with non-zero coefficients. Table7 shows the coefficients of all the covariates estimated from the Lasso regression model. Notably, demographic variables such as the proportion of Aboriginal, Torres Strait Islander, and individuals identifying as both Aboriginal and Torres Strait Islander, along with retail indicators such as liquor store count, pharmacy count, and library presence, were positively associated with EEA counts. In contrast, socio-economic variables, including the

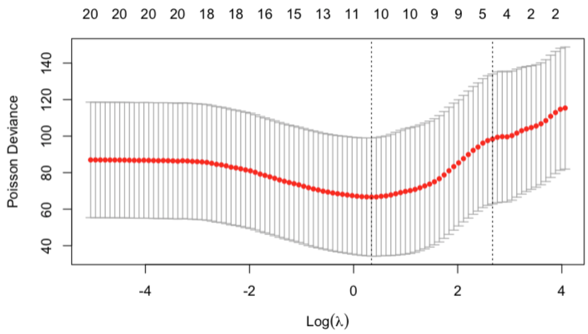


Figure 9: Lasso regression feature selection

proportion of individuals who attended school up to Year 8, average income, and doing blue-collar occupations, showed negative associations with EEA.

Variable	**Coefficient**
Intercept	1.9932
Non-Indigenous	.
Aboriginal	0.5280
Torres Strait Islander	0.1998
Both Aboriginal and Torres Strait Islander	0.0128
Age: 18 and below	.
Age: 18 to 64	.
Age: 65 and over	.
School Years 9–12 Attended	.
Year 8 or Below	-0.1972
School Not Attended	.
Negative Income	.
Nil Income	.
Average Income	-0.5678
Blue Collar Occupation	-0.7464
White Collar Occupation	.
School Count	.
Liquor Outlet Count	0.1996
Pharmacy Count	0.5331
Library Count	0.0354
GP Count	-0.3883

Table 7: Covariates Coefficients from Lasso regression

4.3.2 Spatial regression analysis using CAR model

A series of spatial models were fitted using Lasso-selected covariates potentially associated with EEAs. Four models were evaluated, and model performance was compared based on the marginal log-likelihood values, with a higher (i.e., less negative) marginal log-likelihood indicating a better fit. Model 4, with a marginal log-likelihood of -504.29, was the preferred model. The model assumed a negative binomial distribution and accounted for spatial effects. Table 8 summarises all the models evaluated and their respective marginal log-likelihoods.

Model	Distribution	Spatial Effects	MLL
Model 1	Poisson	No	-563.01
Model 2	Negative Binomial	No	-521.58
Model 3	Poisson	Yes	-544.37
Model 4	Negative Binomial	Yes	$\mathbf{-504.29}$

Table 8: Model Comparison Based on Marginal Log-Likelihood (MLL)

Table 10 illustrates the estimation of the model parameters and overall effect of each parameter on the EEA counts with 95%credible interval (2.5 percentile and 97.5percentile) displayed. All covariates retained from the Lasso selection step whose 95% credible intervals did not overlap zero were considered statistically significant. EEA count was found to be positively associated with the proportion of Aboriginal population (posterior mean = 0.428, 95% CI: 0.089–0.601) and the number of liquor outlets (posterior mean = 0.187, 95% CI: 0.002–0.372).These results indicate that the number of EEAs increased with an increase in the number of Aboriginal residents and liquor stores in a postcode. Although covariates for individuals identifying as both Aboriginal and Torres Strait, individuals who attended school up to year 8, and the number of pharmaceutical facilities were positively associated with EEA, these were not significant, as the 95% credible intervals overlapped with zero.

EEAs were significantly and negatively associated with average income, with a mean posterior value of -1.066 and a 95% credible interval of (-2.002, -0.110), and with blue collar occupation having a mean posterior value of (-0.591,95% CI (-0.956, -0.226)). This means that a one-unit increase in average income and blue collar occupation is likely to lead to a decrease in the number of EEAs in a postal region. Although other covariates, such as individuals identifying as Torres Strait, library count, and GP counts, were negatively associated, their effects were not significant.

Variable	Mean	SD	2.5%	50%	97.5%	Mode	KLD
Intercept	4.407	3.224	-1.933	4.408	10.740	4.408	0
Aboriginal	0.428	0.089	0.253	0.429	0.601	0.429	0
Torres Strait	-0.010	0.096	-0.200	-0.010	0.178	-0.010	0
Aboriginal + Torres Strait	0.086	0.111	-0.131	0.086	0.305	0.086	0
Year 8 or Below	0.024	0.172	-0.311	0.024	0.365	0.024	0
Average Income	-1.066	0.486	-2.022	-1.066	-0.110	-1.066	0
Blue Collar Occupation	-0.591	0.186	-0.956	-0.591	-0.226	-0.591	0
Liquor Outlet Count	0.187	0.094	0.002	0.187	0.372	0.187	0
Pharmacy Count	0.376	0.233	-0.081	0.376	0.834	0.376	0
Library Count	-0.094	0.256	-0.595	-0.094	0.411	-0.094	0
GP Count	-0.278	0.278	-0.825	-0.277	0.266	-0.277	0

Table 9: Bayesian Estimation of social-economic, environmental and retail factors of EEA occurrences in far north Queensland.

The model can thus be interpreted as shown in the Equation 4.

$$\log(\lambda_i) = \beta_0 + \beta_1 \cdot \text{Aboriginal}_i + \beta_2 \cdot \text{AvgIncome}_i + \beta_3 \cdot \text{BlueCollar}_i + \beta_4 \cdot \text{LiquorCount}_i + s_i \quad (4)$$

Where:

$\lambda_i \rightarrow$ Relative risk of EEA in area i

Variable	Mean	SD	2.5%	50%	97.5%	Mode	KLD
Intercept	**4.407**	**3.224**	**-1.933**	**4.408**	**10.740**	**4.408**	**0**
Aboriginal	**0.428**	**0.089**	**0.253**	**0.429**	**0.601**	**0.429**	**0**
Torres Strait	-0.010	0.096	-0.200	-0.010	0.178	-0.010	0
Aboriginal + Torres Strait	0.086	0.111	-0.131	0.086	0.305	0.086	0
Year 8 or Below	0.024	0.172	-0.311	0.024	0.365	0.024	0
Average Income	**-1.066**	**0.486**	**-2.022**	**-1.066**	**-0.110**	**-1.066**	**0**
Blue Collar Occupation	**-0.591**	**0.186**	**-0.956**	**-0.591**	**-0.226**	**-0.591**	**0**
Liquor Outlet Count	**0.187**	**0.094**	**0.002**	**0.187**	**0.372**	**0.187**	**0**
Pharmacy Count	0.376	0.233	-0.081	0.376	0.834	0.376	0
Library Count	-0.094	0.256	-0.595	-0.094	0.411	-0.094	0
GP Count	-0.278	0.278	-0.825	-0.277	0.266	-0.277	0

Table 10: Bayesian Estimation of social-economic, environmental and retail factors of EEA occurrences in far north Queensland.

β_0 $\rightarrow$ Intercept term

Aboriginal$_i$ $\rightarrow$ Proportion of Aboriginal population in area i

AvgIncome$_i$ $\rightarrow$ Average income in area i

BlueCollar$_i$ $\rightarrow$ Proportion of blue-collar workers in area i

LiquorCount$_i$ $\rightarrow$ Number of liquor outlets in area i

s_i $\rightarrow$ Spatially structured random effect (CAR)

Figure 10 shows the a relative risk choropleth map plot EEA count in each postal region. Three postal regions Cairns, Townsville and Mackay all had relatively higher risk compared to other regions. These postal regions with relatively higher risks were evident during the exploratory analysis.

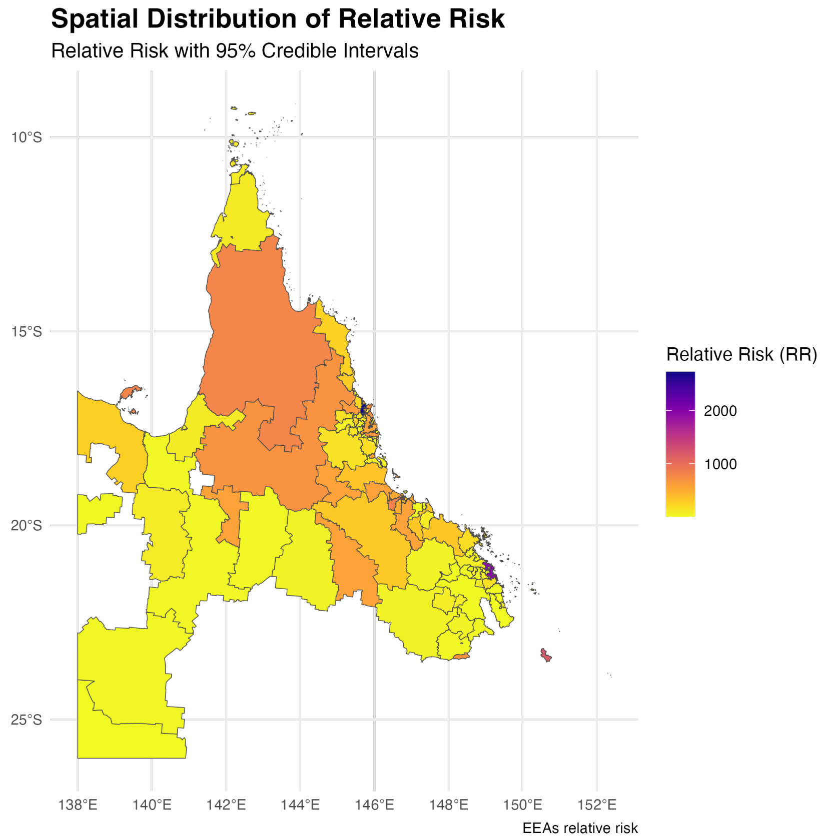


Figure 10: Caption

5 Project Management

5.1 Project Timeline

Figure 11 shows the Gannt chart with timeline for the project. The first two weeks were spent on data processing and aggregation. A significant amount of time was spent on statistical modelling and writing the report.

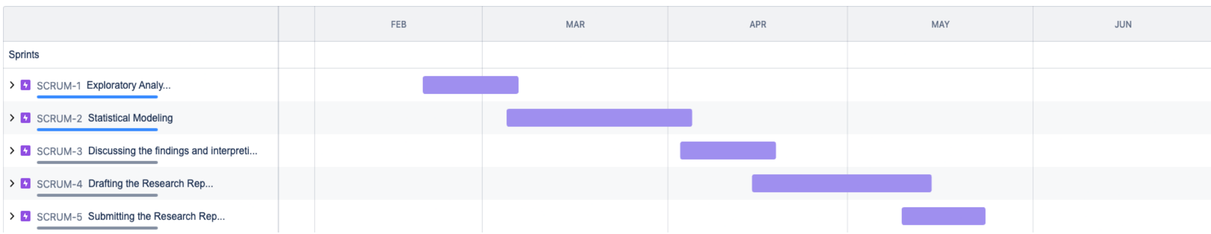


Figure 11: Project timeline

5.2 Management Tools

5.2.1 Agile

- **Notion** was used to for managing all the weekly tasks including weekly progress, assigning and organising weekly submission. Refer Appendix 8.5 for the weekly progress.
- **Git and Bitbucket** were used to track different versions of the project both locally and in the cloud.

- **Microsoft Team Meeting** was used for organising weekly meeting with the supervisor from 10:00 to 11:00 am every Monday.

5.2.2 Data Processing and Analysis

- **Python** was employed for data manipulation and exploratory analysis, utilising robust libraries such as `pandas` and `NumPy` to handle and process large datasets efficiently.
- **R** was primarily used for advanced statistical modelling and spatial analysis, taking advantage of its comprehensive ecosystem of statistical packages.

5.2.3 Software Tools

- **Microsoft Excel** was used during the initial stages of data preparation, particularly for basic data cleaning operations such as removing redundant headers, reformatting, and ensuring structural consistency of tabular data from ABS.

5.2.4 Integrated Development Environments (IDEs)

- **Visual Studio Code (VS Code)**: It was used as code editor of choice for initial phase of the project involving data cleaning, formatting and aggregation.
- **Jupyter Notebook** he primary environment for coding, documentation, and visualisation.
- **RStudio** was utilised for executing R scripts and statistical modelling.

6 Conclusion

The study explored the uncertainties surrounding mental health challenges and the spatial dependencies that emerge when considering various confounding variables. Our findings indicate that mental health-related presentations at emergency departments under Emergency Examination Authority (EEA) are spatially autocorrelated, with neighbouring regions exhibiting similar incidence levels. This may be attributed to shared environmental conditions, socio-economic disadvantages, and community-level infrastructures that influence the prevalence of EEAs.

The distribution of EEAs was not spatially random; rather, it exhibited significant clustering, with geographically proximate regions tending to have similar EEA counts. This clustering was most evident in three key districts: Cairns, Townsville, and Mackay. These regions not only reported the highest levels of EEAs but also shared elevated values for several socio-demographic and retail-related covariates. This spatial pattern suggests that underlying contextual factors, such as demographic factors and access to facilities, may play a significant role EEA observations.

One of the most notable associations uncovered was between EEA clustering and the concentration of liquor outlets. Regions with a higher number of liquor stores consistently demonstrated elevated EEA counts. This finding aligns with previous research by Coomber [37], which observed that alcohol-related incidents tend to cluster in proximity to liquor stores and follow temporal patterns. Furthermore, this support our initial hypothesis that drug-related and substance abuse are among the key reasons for EEA occurrences and their repeated incidence. The statistical model also indicated a strong positive association between number of liquor outlets and EEAs.

The CAR model with spatial interaction estimation also revealed a strong association between demographic factors, particularly the proportion of the Aboriginal population and the incidence of EEAs. This underscores the disparities and access challenges faced by communities in regional Queensland, where limited mental health support may contribute to higher EEA occurrences. In contrast, the statistical model demonstrated a negative association between socioeconomic factors such as average income and EEA counts. Specifically, regions with higher concentrations of individuals employed in blue-collar occupations and potentially earning lower average incomes exhibited higher EEA incidences. This may reflect the mental health challenges associated with occupational stress and socioeconomic disadvantage.

This study also has several limitations. Reliable and comprehensive data on EEAs, as well as on socioeconomic, retail, and environmental variables, are not readily available. The datasets used in this study required extensive preprocessing and transformation to curate a version suitable for analysis. Another important limitation is the lack of temporal analysis. Although the EEA records span from 2008 to 2021, this study did not explore temporal variations in the data. Understanding the temporal dynamics of EEAs could provide deeper insights into their relationship with confounding factors and help identify trends and changes over time.

EEA is influenced by demographic factors, accessibility to retail facilities, and socio-economic factors. The study indicated clear patterns of spatial variation in EEAs and the association with confounding factors. Areas with a higher concentration of liquor outlets and higher proportions of Aboriginal residents are likely to experience higher EEAs. Similarly, regions with a greater proportion of residents in lower income groups and those engaged in blue-collar occupations are also likely to exhibit higher EEA occurrences.

This study would contribute significantly to the field of mental health research and opens avenues for more efficient and data-driven policy-making in addressing mental health issues. The research lays a novel and foundational framework for data collection, processing, and assimilation from different sources into a single database registry for mental health, and statistical modelling using the CAR model, which none of the previous studies have implemented. The insights from the study would enable researchers and policy makers to design intervention strategies for emergent mental health conditions.

7 Reference

[1] Australian Institute of Health and Welfare, "Mental health expenditure," 2023, accessed: 2024-10-09. [Online]. Available: https://www.aihw.gov.au/mental-health/topic-areas/expenditure

[2] S. Judkins, D. Fatovich, N. Ballenden, and H. Maher, "Mental health patients in emergency departments are suffering: the national failure and shame of the current system," *Australas Psychiatry*, vol. 27, no. 6, pp. 615–617, Dec. 2019.

[3] B. E. Kavanagh, K. B. Corney, H. Beks, L. J. Williams, S. E. Quirk, and V. L. Versace, "A scoping review of the barriers and facilitators to accessing and utilising mental health services across regional, rural, and remote australia," *BMC Health Services Research*, vol. 23, no. 1, p. 1060, Oct. 2023.

[4] J. Alcock, J. C. Oam, J. Ranse, and R. Wardrop, "Characteristics and outcomes of emergency department presentations brought in by police with and without an emergency examination authority: A state-wide cohort study," *Australasian Emergency Care*, vol. 27, no. 3, pp. 185–191, Sep. 2024.

[5] Queensland Government, "Public health act 2005," 2024, accessed: 2024-08-28. [Online]. Available: https://www.legislation.qld.gov.au/view/html/inforce/current/act-2005-048

[6] A. R. Clough *et al.*, "Emergency examination authorities in queensland, australia," *Emerg Medicine Australasia*, vol. 35, no. 5, pp. 731–738, Oct. 2023.

[7] S. Das, J. Catterall, R. Stone, and A. R. Clough, ""the reasons you believe …": An exploratory study of text driven evidence gathering and prediction from first responder records justifying state authorised intervention for mental health episodes," *Computer Methods and Programs in Biomedicine*, vol. 254, p. 108257, Sep. 2024.

[8] T. Meehan and T. Stedman, "Trends in the use of emergency examination orders in queensland since the implementation of the mental health intervention project," *Australas Psychiatry*, vol. 20, no. 4, pp. 287–290, Aug. 2012.

[9] A. R. Clough *et al.*, "Recent amendments to queensland legislation make mental health presentations to hospital emergency departments more difficult to scrutinise," *Emerg Medicine Australasia*, vol. 34, no. 1, pp. 130–133, Feb. 2022.

[10] P. G. Miller *et al.*, "Queensland alcohol-related violence and night-time economy monitoring (quantem): Rationale and overview," *Drug and Alcohol Review*, vol. 40, no. 5, pp. 693–697, Jul. 2021.

[11] E. Amsalu *et al.*, "Spatial-temporal analysis of tuberculosis in the geriatric population of china: An analysis based on the bayesian conditional autoregressive model," *Archives of Gerontology and Geriatrics*, vol. 83, pp. 328–337, Jul. 2019.

[12] P. Moraga. (2024) Chapter 8 spatial autocorrelation — spatial statistics for data science: Theory and practice with r. Accessed: Aug. 14, 2024. [Online]. Available: https://www.paulamoraga.com/book-spatial/spatial-autocorrelation.html

[13] Australian Bureau of Statistics, "Snapshot of queensland," 2024, accessed: 2024-09-18. [Online]. Available: https://www.abs.gov.au

[14] Queensland Health, "Data management standard (qh-imp-279-4:2023)," Health Informatics Services, eHealth Queensland, Deputy Director-General, Sep. 2023, accessed: May 15, 2025. [Online]. Available: https://www.health.qld.gov.au/__data/assets/pdf_file/0011/1210142/qh-imp-279-4.pdf

[15] ——, "Data management policy (qh-pol-279:2014)," Health Informatics Services, eHealth Queensland, Deputy Director-General, Nov. 2014, accessed: May 15, 2025. [Online]. Available: https://www.health.qld.gov.au/__data/assets/pdf_file/0025/396052/qh-pol-279.pdf

[16] Queensland Government, Department of Justice and Attorney-General, "Liquor and gaming services," 2024, accessed: 2024-08-23. [Online]. Available: https://www.justice.qld.gov.au/about-us/services/liquor-gaming

[17] Australian Health Practitioner Regulation Agency, "Regulating australia's health practitioners," 2024, accessed: 2024-9-18. [Online]. Available: https://www.ahpra.gov.au/

[18] Queensland Government, "Open data portal," 2024, accessed: 2024-9-19. [Online]. Available: https://www.data.qld.gov.au/

[19] Queensland Government, Department of Health, "Pharmacy business ownership administration system," 2024, accessed: 2024-10-16. [Online]. Available: https://www.pboas.health.qld.gov.au/

[20] Australian Commission on Safety and Quality in Health Care, "Australian commission on safety and quality in health care," Australian Government Department of Health and Aged Care, May 2025, accessed: May 22, 2025. [Online]. Available: https://www.safetyandquality.gov.au/

[21] Selenium HQ, "Selenium," 2024, accessed: 2024-10-16. [Online]. Available: https://www.selenium.dev/

[22] Queensland Health, "Rti 6029/24 – register of pharmacy businesses in queensland," Right to Information Release, Queensland Government, May 2024, accessed: April 20, 2025. [Online]. Available: https://www.health.qld.gov.au/__data/assets/pdf_file/0025/1339144/RTI-6029-Pharmacy-Businesses-in-Queensland.PDF

[23] Microsoft Corporation, "Microsoft excel," 2024, accessed: 2024-10-16. [Online]. Available: https://www.microsoft.com/en-us/microsoft-365/excel

[24] X. Hu, S. Kaplan, and R. S. Dalal, "An examination of blue- versus white-collar workers' conceptualizations of job satisfaction facets," *Journal of Vocational Behavior*, vol. 76, no. 2, pp. 317–325, Apr. 2010.

[25] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2021. [Online]. Available: https://www.R-project.org/

[26] R. Bivand and D. W. S. Wong, "Comparing implementations of global and local indicators of spatial association," *TEST*, vol. 27, no. 3, pp. 716–748, 2018.

[27] Roger Bivand, “R packages for analyzing spatial data: A comparative case study with areal data,” *Geographical Analysis*, vol. 54, no. 3, pp. 488–518, 2022.

[28] R. S. Bivand, E. Pebesma, and V. Gómez-Rubio, *Applied spatial data analysis with R, Second edition*. Springer, NY, 2013. [Online]. Available: https://asdar-book.org/

[29] E. Pebesma and R. S. Bivand, *Spatial Data Science With Applications in R*. Chapman Hall, 2023. [Online]. Available: https://r-spatial.org/book/

[30] H. Rue, S. Martino, and N. Chopin, “Approximate bayesian inference for latent gaussian models using integrated nested laplace approximations (with discussion),” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 71, pp. 319–392, 2009.

[31] T. G. Martins, D. Simpson, F. Lindgren, and H. Rue, “Bayesian computing with inla: New features,” *Computational Statistics and Data Analysis*, vol. 67, pp. 68–83, 2013.

[32] F. Lindgren *et al.*, “An explicit link between gaussian fields and gaussian markov random fields: the stochastic partial differential equation approach [with discussion],” *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, vol. 73, no. 4, pp. 423–498, 2011.

[33] F. Lindgren and H. Rue, “Bayesian spatial modelling with r-inla,” *Journal of Statistical Software*, vol. 63, no. 19, pp. 1–25, 2015. [Online]. Available: http://www.jstatsoft.org/v63/i19/

[34] P. A. P. Moran, “The interpretation of statistical maps,” *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 10, no. 2, pp. 243–251, Jul. 1948.

[35] Y. Wang and K. M. Kockelman, “A poisson-lognormal conditional-autoregressive model for multivariate spatial analysis of pedestrian crash counts across neighborhoods,” *Accident Analysis & Prevention*, vol. 60, pp. 71–84, Nov. 2013.

[36] G. Wang, “Laplace approximation for conditional autoregressive models for spatial data of diseases,” *MethodsX*, vol. 9, p. 101872, 2022.

[37] K. Coomber *et al.*, “The impact of liquor legislation changes on police-recorded serious assault in queensland, australia,” *Drug and Alcohol Review*, vol. 40, no. 5, pp. 717–727, Jul. 2021.

8 Appendix

8.1 Appendix A1: Map of Queensland

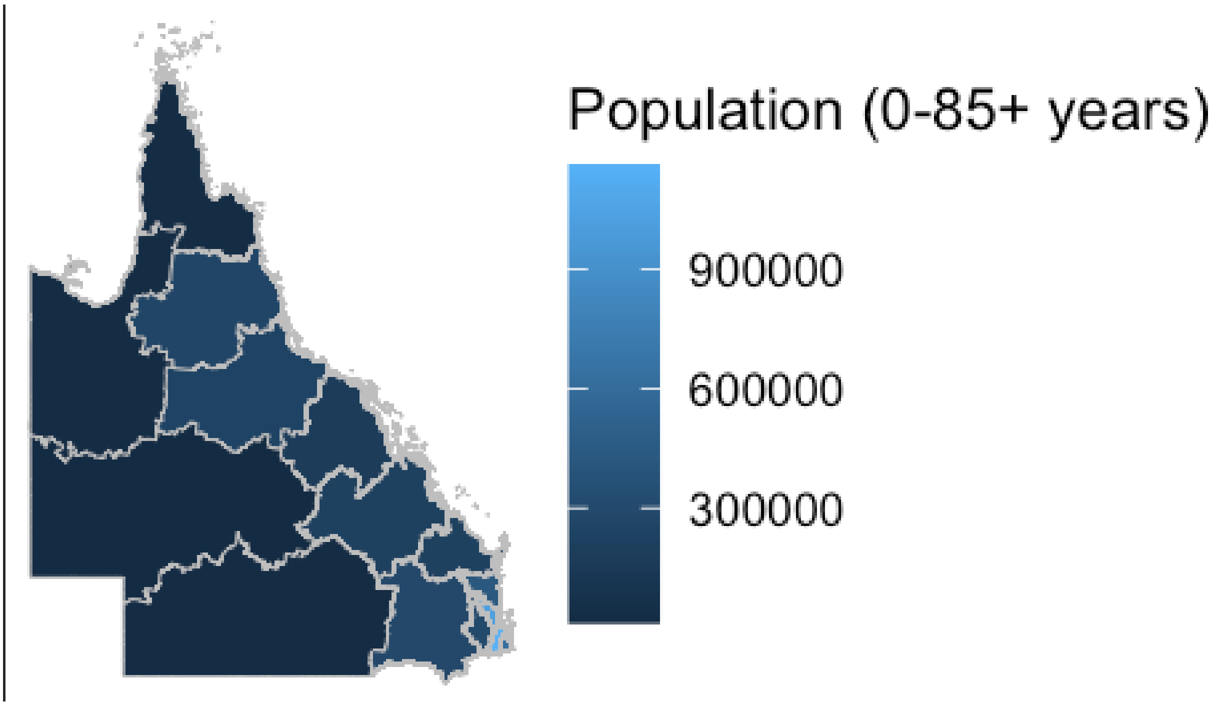


Figure 12: Map of Queensland

8.2 Appendix A: Email Template

Dear Team,

I hope this message finds you well.

I am writing to request access to specific data maintained by your agency for the purpose of conducting academic research. My name is Nidup Dorji, and I am a student at Curtin University. I am currently studying an epidemiological academic research part of my degree program. The project considers various socio-economic factors and well—being metrics in far north Queensland under the guidance of Professor Alan Clough of James Cook University and Dr Sourav Das at Curtin University.

To facilitate our study, we kindly request access to the following data:

1. Number of liquor stores in each postcode, including trading hours.

2. Data to be provided in CSV format.

If there are any procedures we need to follow to obtain the data, please let us know, and we will follow up promptly.

Thank you for considering our request. We look forward to your response.

Warm regards,

Nidup Dorji
Student
Curtin University

8.3 Appendix C: Python Scrapper Script

The following code implements a web scraper for general practices:

```python
from selenium import webdriver
from selenium.webdriver.chrome.service import Service
from selenium.webdriver.common.by import By
from selenium.webdriver.support.ui import WebDriverWait
from selenium.webdriver.support import expected_conditions as EC
from selenium.webdriver.chrome.options import Options
from webdriver_manager.chrome import ChromeDriverManager
from selenium.common.exceptions import WebDriverException
import time

def gp_scraper(url, user_agent):
    # Set up Chrome options
    chrome_options = Options()
    chrome_options.add_argument("--headless")  # Run without GUI
    chrome_options.add_argument(f"user-agent={user_agent}")

    # Set up Chrome driver
    service = Service(ChromeDriverManager().install())
    driver = webdriver.Chrome(service=service, options=chrome_options)

    try:
        # Navigate to the URL
        driver.get(url)
        # Allow the page to load
        time.sleep(3)  # Adjust sleep time as needed

        # Check the page title or a specific element to confirm accessibility
        if "403" in driver.title or "not found" in driver.title.lower():
            print("Access forbidden or page not found.")
        else:
            # Wait for the results page to load (modify condition as needed)
            WebDriverWait(driver, 10).until(EC.presence_of_element_located((By.CLASS_NAME, "list")))  # Adjust class name

            # Get the outerHTML of the entire page or a specific element
            itemlists = driver.find_elements(By.CLASS_NAME, "list-item")  # Get outerHTML of body
            gp = []
            for i in range(min(14, len(itemlists))):  # Ensure we don't go out of bounds
                # Dictionary to store each gp
                gp_detail = dict()
                try:
```

```
                    category = itemlists[i].find_element(By.CSS_SELECTOR, "
                        ↪ div.item-category > div.ng-binding").text
                    gp_detail["category"] = category
                except WebDriverException:
                    continue

                try:
                    # Get the details wrapper
                    details_wrapper = itemlists[i].find_element(By.
                        ↪ CSS_SELECTOR, "div.detailsWrapper")

                    name = details_wrapper.find_element(By.CSS_SELECTOR, "a
                        ↪ .item-title > h2").text
                    gp_detail["name"] = name

                    # Get the address from the details wrapper
                    address = details_wrapper.find_element(By.CSS_SELECTOR,
                        ↪  "div.item-address .ng-binding").text
                    gp_detail["address"] = address
                except WebDriverException:
                    continue

                gp.append(gp_detail)
            return gp

    except WebDriverException as e:
        print("WebDriverException occurred:", e)
    finally:
        driver.quit()
```

For more information, visit the following link: https://www.truelocal.com.au/search/general-practice-and-mental-health/queensland?page=1.

8.4 Appendix D: Merge Validation

```
#create test class
class DataMergeTest():

    #should take two parameters (unmerged data, and merged data)
    def __init__(self, unmerged_data, merged_data):
        self.unmerged_data = pd.read_csv(unmerged_data, index_col=None)
        self.merged_data = pd.read_csv(merged_data, index_col=None)
        # print(self.unmerged_data)

    #test function
    def test_merge(self):
        #select random 5 rows from the merge data
        sample = self.merged_data.sample(5)
        random_postcodes = sample['postcode']

        unmerged_rows = self.unmerged_data[self.unmerged_data['postcode'].
            ↪ isin(random_postcodes)]
        # print(unmerged_rows)
        merged_rows = self.merged_data[self.merged_data['postcode'].isin(
            ↪ random_postcodes)]
        # print(merged_rows)
        #since the merged data has lots of columns, we need to select one
            ↪ that is only in unmerged data
```

```
        select_columns = list(unmerged_rows.columns)
        merged_rows_selected_columns = merged_rows[select_columns]
        # print(merged_rows_selected_columns)
        # print(unmerged_rows)

        #check if the selected columns from merged data and unmerged data are equal
        for row in range(unmerged_rows.shape[0]):
            is_equal = unmerged_rows.iloc[row].equals(merged_rows_selected_columns.iloc[row])
            print(f"Values in rows {row} for both merge and unmerged data: {is_equal}")
            if not is_equal:
                print("Differences:")
                comparison = unmerged_rows.iloc[row] == merged_rows_selected_columns.iloc[row]
                print(comparison[~comparison])
            assert is_equal
```

8.5 Appendix E - Weekly Progress Report

The link to the weekly progress. Weekly Progress Reports