COMP5070_SP2_2023
Exam
Data Science Job Market in Australia

Enna H

Introduction

Amidst the swift expansion of the Data Science field and its proliferating demand across numerous sectors, gaining an insightful understanding of the job market's dynamics is pivotal for both employers and job seekers to remain competitive and make informed decisions.

This report conducts an extensive exploration of the Data Science job market within Australia, utilizing data obtained from Seek.com.au, Australia's leading job search website. Data encompassing 540 search results was acquired through web scraping, with the majority of these postings created in the 30 days preceding June 24, 2023. It is important to note that some featured postings may be displayed on the Seek site for longer than 30 days. The analysis pays attention to crucial aspects such as salary distribution, job market diversity across states, prevalent job titles and industries, and a time-based analysis of job postings. The overarching objective of this analysis is to generate valuable insights that will aid job seekers in identifying promising opportunities and employers in formulating effective recruitment strategies.

Exploratory Data Analysis

The exploratory data analysis section delves into the dataset obtained from Seek.com.au, focusing on key statistical findings related to annual salaries in the Data Science domain. By examining summary statistics such as median salary, standard deviation, maximum salary, this section offers a comprehensive overview of the salary landscape. Additionally, visualizations in the form of histograms provide a clear understanding of the salary distribution, aiding job seekers and employers in grasping the prevailing salary levels within the market.

Table 1: Summary Statistics for Annual Salary in Seek Data Scientist Dataset

salary_median	salary_sd	salary_max
112682.00	28738.37	180000.00

Not all job postings include the annual wage. Table 1 shows the summary statistics for the annual pay in the Seek Data Scientist ad that includes the compensation in the description. With a standard deviation of $28738.37, the median pay is $11,2682. The highest recorded compensation is $180,000.

Table 2: Number of unique values for each feature

	Unique_Counts
job_title	354
area	23
industry	18
state	8

Table 2 presents the number of unique values for each feature in the dataset. There are 354 unique job titles, 23 unique areas, 18 unique industries, and 8 unique states represented in the dataset.

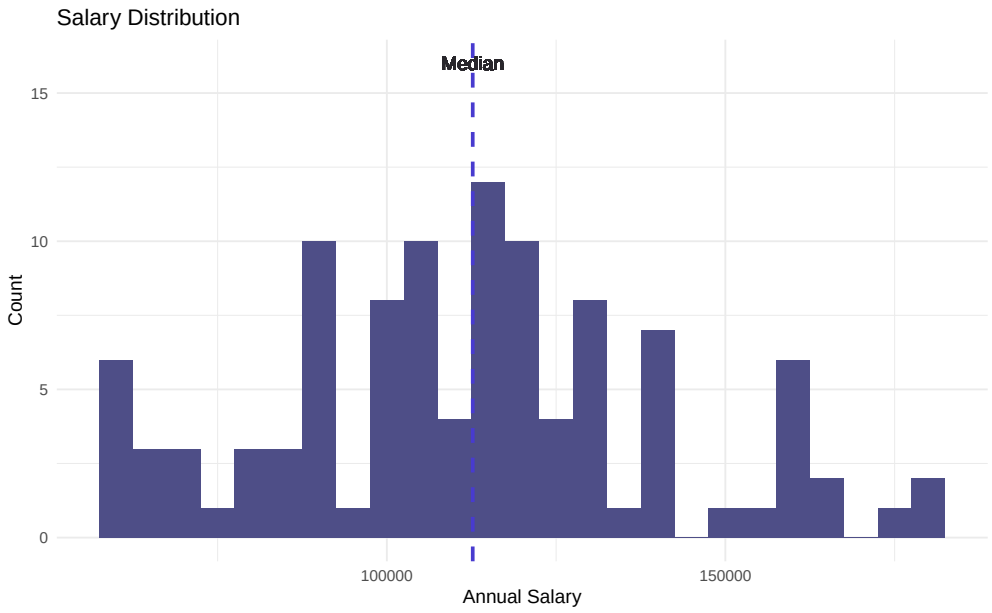

Figure 1. Salary Distribution in Seek Data Scientist Dataset.

Figure 1 illustrates the distribution of salaries in the Seek Data Scientist dataset. The histogram provides an overview of the salary ranges for data science positions. Further analysis of the distribution can help job seekers and employers understand the prevailing salary levels in the market.

The EDA of the Seek.com.au dataset has provided us with several key insights about the landscape of salaries in the Data Science domain. The median annual salary for job postings that include a salary is approximately $112,682, demonstrating that data science is a well-compensated field overall. The highest salary recorded is $180,000, indicating that there are potentially very lucrative opportunities available for highly skilled or experienced professionals.

The large standard deviation of approximately $28,738 in salaries underscores the variation in the dataset. This could be due to factors such as job level (junior vs. senior roles), specific technical skills required, or even geographical location. Therefore, it's important for job seekers and employers to bear in mind that salaries can differ significantly based on these parameters.

The dataset comprises a large variety of job titles, areas, industries, and states. This diversity reflects the dynamic and evolving nature of the data science field. The fact that there are 354 unique job titles suggests that there may be a broad array of roles that fall under the data science umbrella, each potentially carrying different salary expectations.

The histogram visualization offers a clear representation of salary distribution, showing that while there are outliers, a significant proportion of salaries are clustered around the median. This can be a useful reference for both job seekers and employers to understand the standard market rate for data science professionals.

While the EDA has offered a general understanding of the salary scenario, it's important to consider that the data is specific to Seek.com.au, which may not encompass all job postings in the field of Data Science. Additionally, the EDA doesn't account for certain details such as years of experience, educational qualifications, or specific skills, which could potentially have a significant impact on the salary. Hence, this analysis provides a starting point, but a more detailed analysis considering these factors could provide further depth to our understanding.

Job Market Distribution across States

This section investigates the distribution of Data Science job positions across different states in Australia. Through the utilization of bar charts, it presents a visual representation of the number of available positions in each state. Furthermore, employing boxplots, this section examines the salary distribution per state, aiming to identify states that offer higher salaries for Data Scientist roles. By combining visual and tabular data, this analysis helps job seekers and employers comprehend the regional dynamics of the job market and make informed decisions accordingly.

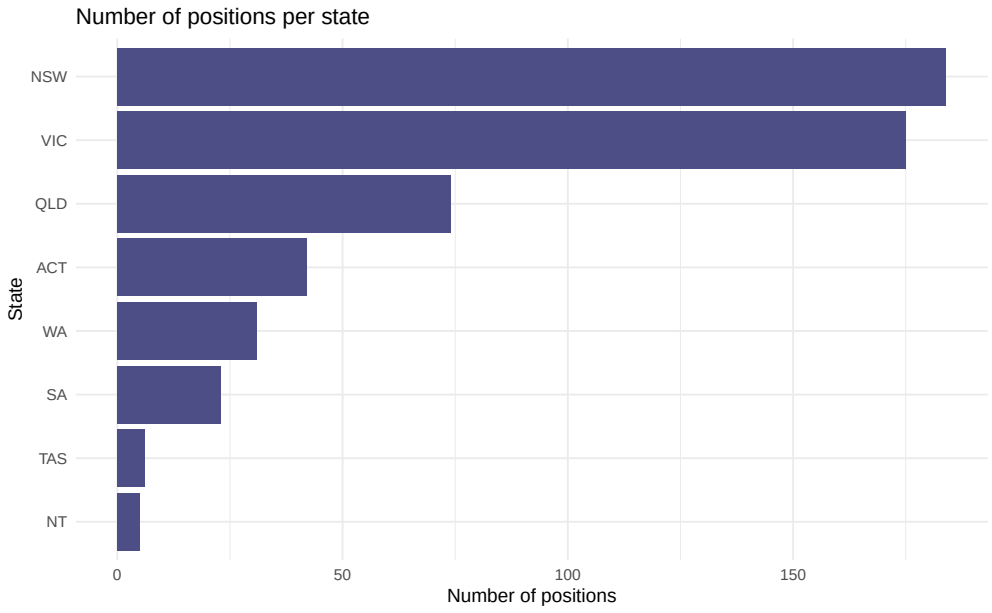

Figure2: Number of positions per state.

Now that we have visualized the number of positions available in each state, let's compare the salary distributions across states. This will provide insights on whether certain states tend to offer higher salaries for Data Scientist roles.

Let's use a boxplot to visualize the salary distribution per state: Boxplot for salary_annual by state or industry: This can give you a sense of the salary distribution per state or industry.

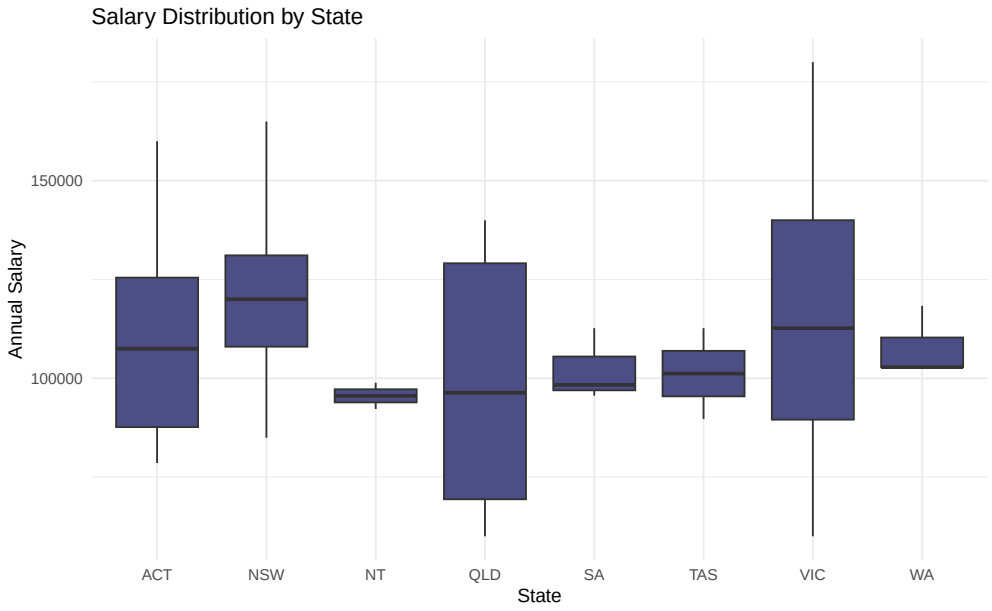

Figure 3: Salary distribution by state.

Table 3: the number of positions per state and median salary per state

state	n	median_salary
ACT	42	107475
NSW	184	120000
NT	5	95548
QLD	74	96341
SA	23	98313
TAS	6	101181
VIC	175	112682
WA	31	102808

The analysis reveals that New South Wales (NSW) and Victoria (VIC) dominate the Data Science job market, with 184 and 175 positions respectively. Despite being considerably smaller, the Australian Capital Territory (ACT) also offers a substantial number of positions, housing 42 opportunities. Conversely, Tasmania (TAS) and the Northern Territory (NT) proffer the fewest opportunities, with 6 and 5 positions respectively.

The analysis of salary distributions reveals that NSW offers the highest median salary at AUD 120,000, followed by VIC with AUD 112,682 and ACT with AUD 107,475. The lowest median salary is recorded in NT, amounting to AUD 95,548. Thus, NSW and VIC emerge as the most attractive destinations for data scientists, with a plethora of job opportunities and competitive median salaries. On the other hand, TAS and NT, while providing fewer opportunities, still offer relatively competitive salaries.

From a business standpoint, the analysis provides valuable insights into the geographical distribution of data science opportunities and associated salaries in Australia. Evidently, New South Wales (NSW) and Victoria (VIC) are the hotspots for data science jobs, accounting for the majority of the positions. This could be indicative of a more mature technology and business landscape in these states, with numerous companies leveraging data science for their operations.

On the flip side, while Tasmania (TAS) and the Northern Territory (NT) might offer fewer job opportunities, they do present potential growth areas for businesses looking to tap into less competitive markets. As data

science continues to grow, we could see an increase in opportunities in these regions, making them ideal for businesses considering future expansions.

In terms of salaries, NSW, VIC, and the Australian Capital Territory (ACT) lead with higher median salaries. Companies in these states seem to value the role of data scientists, offering competitive remuneration to attract top talent. However, businesses must consider this information in tandem with the cost of living and operation in these states, which could be higher than in regions with lower median salaries.

Furthermore, the discrepancy in salaries across states can influence businesses' talent acquisition strategies. For instance, they could consider remote working policies to tap into talent pools beyond their state borders.

This analysis also signals the importance of data-driven decision-making across various sectors. As data science opportunities grow, it's crucial for businesses not just to invest in hiring skilled professionals but also to ensure they are offering competitive salaries to retain them. Regular benchmarking against market trends, like the analysis presented, can help businesses stay competitive in their compensation strategies.

However, it's worth noting that these trends are subject to change, as the technology industry is dynamic and rapidly evolving. Continuous monitoring of the job market and salaries across states will be crucial to adapt to changes and make informed business decisions.

Job Titles and Industries

The job titles and industries section explores the most common job titles and industries within the Data Science job market. Through the use of bar charts, it provides a visual depiction of the frequency of the top 10 job titles and industries. Furthermore, detailed tables present the specific job titles and industries along with their respective frequencies. The analysis also includes an examination of median salaries by industry, allowing job seekers and employers to identify sectors that offer higher compensation. This section assists in aligning career aspirations and recruitment strategies with the prevailing job titles and industries in the market.

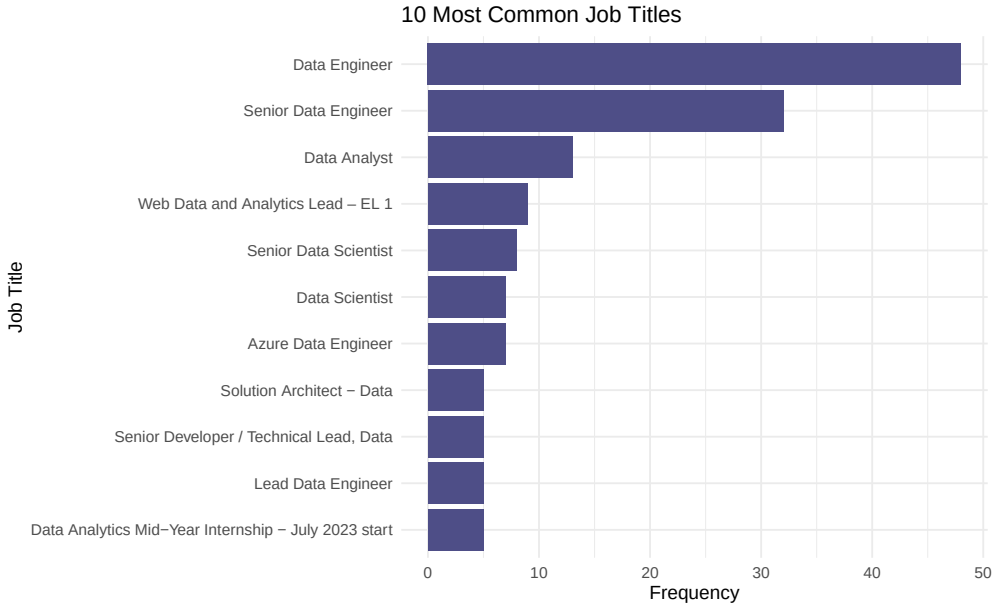

Figure 4: Bar plot displaying the 10 most common job titles in the Seek Data Scientist job postings for the past 30 days.

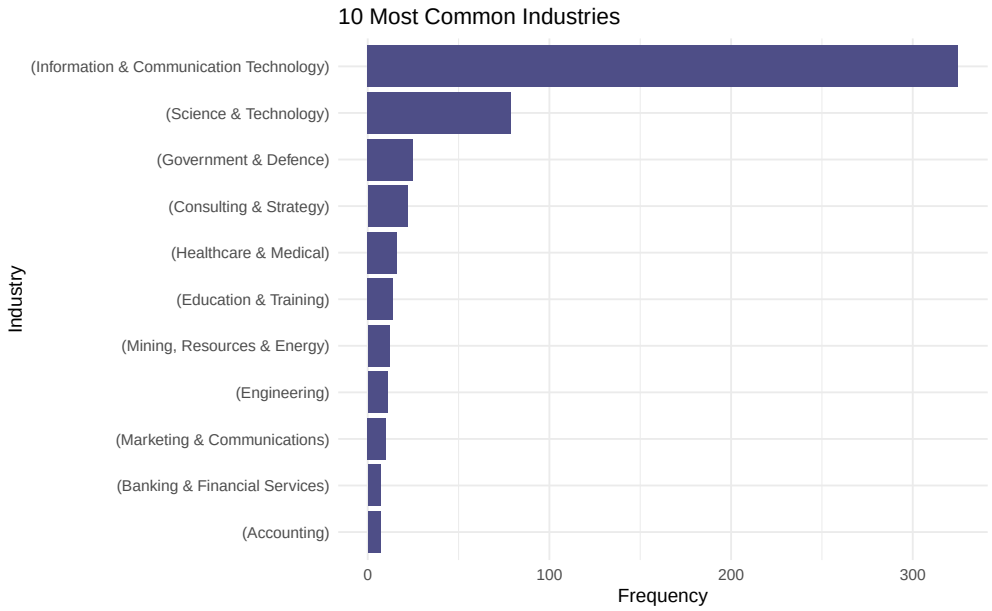

Figure 5: Bar plot displaying the 10 most common industries in the Seek Data Scientist job postings for the past 30 days.

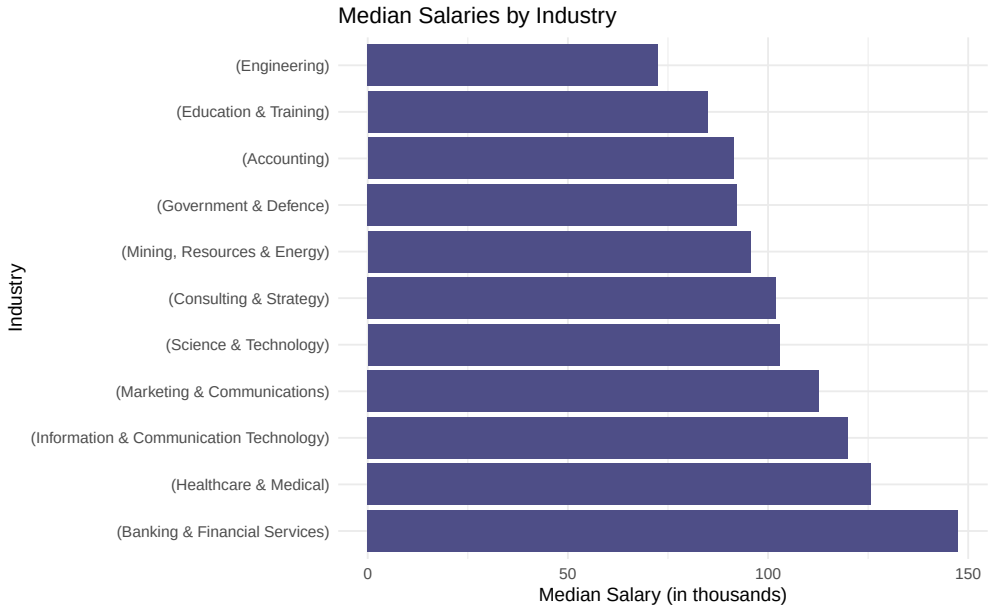

Figure 6: Bar plot displaying the median salaries by industry in the Seek Data Scientist job postings for the past 30 days.

Table 4: 10 Most Common Job Titles

job_title	n
Data Engineer	48
Senior Data Engineer	32
Data Analyst	13

job_title	n
Web Data and Analytics Lead – EL 1	9
Senior Data Scientist	8
Azure Data Engineer	7
Data Scientist	7
Data Analytics Mid-Year Internship - July 2023 start	5
Lead Data Engineer	5
Senior Developer / Technical Lead, Data	5
Solution Architect - Data	5

Table 5: 10 Most Common Industries

industry	n
(Information & Communication Technology)	325
(Science & Technology)	79
(Government & Defence)	25
(Consulting & Strategy)	22
(Healthcare & Medical)	16
(Education & Training)	14
(Mining, Resources & Energy)	12
(Engineering)	11
(Marketing & Communications)	10
(Accounting)	7
(Banking & Financial Services)	7

Table 6: Median Salaries by Industry (in thousands)

industry	median_salary
(Accounting)	91.42
(Banking & Financial Services)	147.50
(Consulting & Strategy)	101.95
(Education & Training)	84.94
(Engineering)	72.50
(Government & Defence)	92.23
(Healthcare & Medical)	125.72
(Information & Communication Technology)	120.00
(Marketing & Communications)	112.68
(Mining, Resources & Energy)	95.59
(Science & Technology)	102.81

When it comes to job titles, “Data Engineer” is the most frequent, followed by “Senior Data Engineer” and “Data Analyst”. The “Information & Communication Technology” industry emerged as the most prevalent industry for data science roles, significantly outnumbering other sectors like “Science & Technology” and “Government & Defence”.

Regarding salary distribution across industries, the “Banking & Financial Services” sector pays the highest median salary (AUD 140,000), followed by “Healthcare & Medical” and “Marketing & Communications” industries. It’s worth noting that these salary metrics are based on job postings and may not necessarily represent the final compensation provided, which could also include benefits and bonuses.

From a business perspective, the analysis of common job titles and industries gives a valuable insight into the job market structure and remuneration practices within different industries.

Understanding the most common job titles such as "Data Engineer," "Senior Data Engineer," and "Data Analyst" can guide businesses in aligning their job descriptions and roles to industry standards. It's apparent that the data engineering field is prominent, indicating a high demand for professionals who can organize and manage large data sets. Businesses, particularly those in the process of building or expanding their data teams, might consider creating more roles in these areas, given their apparent market prominence.

The fact that the "Information & Communication Technology" sector is leading in data science roles is expected, given the sector's reliance on data for decision-making, strategy formulation, and operations. However, the presence of sectors like "Science & Technology," "Government & Defence," and even "Healthcare & Medical" in the top industries shows the increasing adoption of data science across diverse fields. Businesses in these sectors should ensure they stay competitive by incorporating data science into their strategies and operations.

The salary analysis by industry is of particular interest to businesses for multiple reasons. It highlights the "Banking & Financial Services" sector's willingness to invest heavily in data science, indicating their high reliance on data-driven decision-making. It also shows that the "Healthcare & Medical" and "Marketing & Communications" sectors, despite being less represented in job postings, value data science roles highly and are willing to offer competitive salaries.

However, it's important to consider other forms of remuneration, such as benefits and bonuses, which are often not represented in job posting salaries. Companies should take this into account when formulating their compensation packages. Offering competitive salaries could be crucial to attract and retain top talent in these roles.

Moreover, businesses in industries that are currently less represented in data science roles, like "Government & Defence," might consider the potential benefits of expanding their data science capabilities. Given the increasing reliance on data across all sectors, early adoption could offer a significant competitive advantage.

In conclusion, businesses need to consider the insights from these data when formulating strategies related to talent acquisition, compensation packages, and industry trends. Regularly updating this analysis can ensure businesses stay abreast of market trends and can make informed, data-driven decisions.

Time-Based Analysis of Job Postings

This section focuses on the temporal patterns of job postings in the Data Science field. A line chart illustrates the number of job postings over time, offering insights into the fluctuations in job demand. Additionally, a bar chart and a table provide an analysis of job postings based on the day of the week, enabling job seekers and employers to identify patterns and optimize their job search and recruitment efforts accordingly. This time-based analysis aids in understanding the dynamics of job availability and assists in planning effective job search and recruitment strategies.

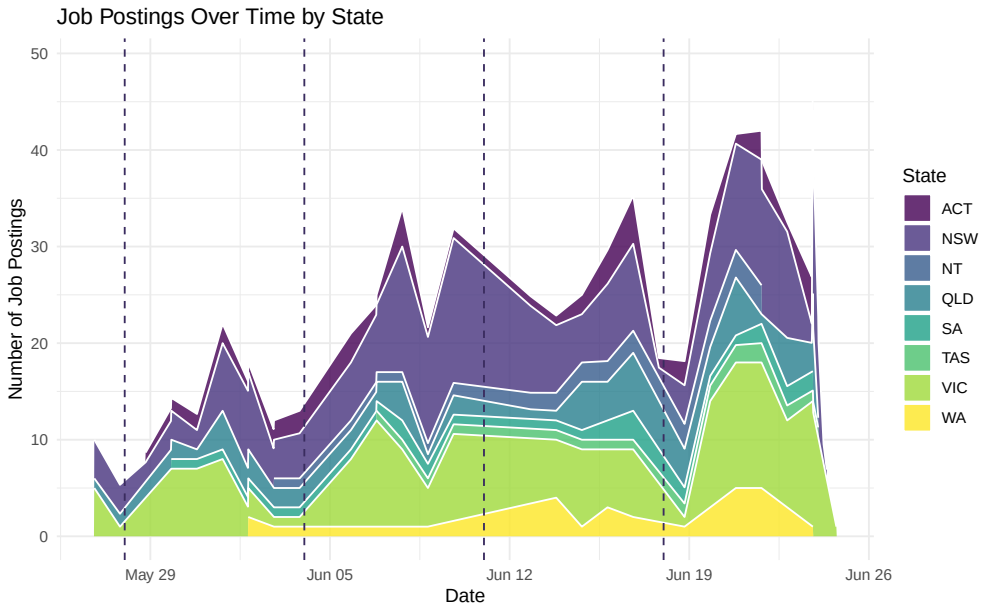

Figure 7: Job Postings Over Time, Stacked by State, with Week Start Indicated by Dashed Lines.

A time series analysis of the job postings reveals significant fluctuations, indicating that the demand for data scientists varies over time. However, a more granular analysis that factors in events such as economic fluctuations or policy changes may be required to provide precise explanations for these variations.

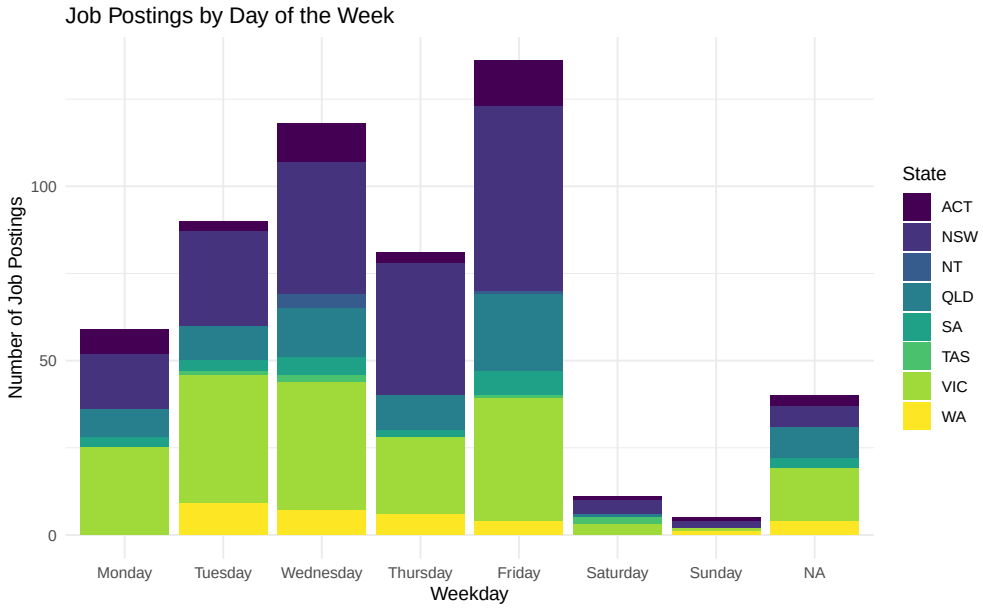

Figure 8: Job Postings by Day of the Week.

Table 7: Job Postings by Day of the Week

weekday	n
Friday	136
Monday	59
Saturday	11
Sunday	5
Thursday	81
Tuesday	90
Wednesday	118
NA	40

Moreover, an analysis of the distribution of job postings across weekdays shows a surge in postings at the start and end of the working week, with the highest number recorded on Fridays. The number of postings drops significantly over the weekend, demonstrating the typical job posting behaviour in line with the conventional work week.

The temporal analysis of job postings provides valuable insights into recruitment patterns and the dynamics of job demand in the data science field. Trend analysis of job postings over time can help businesses understand the fluctuating demand for data science roles, which may be due to multivariate factors such as economic factors or policy changes. This information can guide hiring efforts, such as attracting top talent during periods of increased postings or simplifying the recruitment process during lulls.

The analysis of job postings by the day of the week can help businesses optimize their recruitment strategies. A higher number of postings on Fridays suggests that businesses prefer to advertise new positions towards the end of the work week. Understanding this pattern can help businesses stand out by posting job advertisements on less busy days, reducing competition and increasing visibility. Additionally, job postings drop significantly over the weekend, attracting potential applicants who might be more active in their job search during non-working days.

Strategic planning can benefit from aligning job posting schedules with observed patterns and responding promptly to changing market dynamics. This time-based analysis helps businesses optimize their recruitment efforts, improve visibility of job postings, and align hiring strategies with observed market trends. Overall, this time-based analysis emphasizes the importance of leveraging data to make informed, strategic business decisions.

Conclusions and Recommendations

Australia's Data Science job market is a dynamic and diverse field, providing distinct opportunities and challenges for both job seekers and employers. The analysis has revealed substantial differences across states, job roles, and industries, emphasizing that an in-depth understanding of these variations can significantly contribute to aligning job seekers' career aspirations with suitable opportunities, and assisting employers in maintaining a competitive edge.

Even though states such as NSW and VIC might seem more appealing due to the abundance of opportunities, job seekers should not overlook states like TAS and NT. These regions may offer fewer positions, but the competition is also less intense, making it crucial for candidates to consider a wider geographical range in their job search.

The analysis of temporal patterns in job postings offers valuable insights. Job seekers can harness this information to optimize their job search strategies, while employers can strategically time their job postings to ensure maximum visibility.

Moreover, given the fast-paced evolution of the Data Science field, staying abreast of current trends and changes in the job market is essential. Future studies could further investigate the correlation between job titles, required skills, and salary offerings, providing more detailed insights for stakeholders. Such information could enable job seekers to target high-demand skills and assist employers in designing competitive compensation packages.

In conclusion, this analysis has unearthed key insights about the Australian Data Science job market, highlighting regional and temporal variations, prevalent job titles, and industries. These findings serve as a valuable guide for job seekers and employers in their decision-making process. However, with the ever-evolving landscape, continuous monitoring of market trends and adapting strategies accordingly is crucial. Further research could delve into the intricate relationship between job titles, associated skills, and salaries, thereby providing a more comprehensive understanding of the job market dynamics.