Preliminary Data Analysis

Contents

Introduction

In recent years, there has been a significant increase in demand for data science-related roles, driven by the exponential growth of data across industries worldwide. This surge underscores the urgent need for professionals who can effectively navigate and interpret vast datasets to facilitate informed decision-making. Data science jobs play a pivotal role in this landscape, utilizing their expertise to extract valuable insights from data. (Ramasubbareddy, 2019)

The project aims to explore job opportunities in Data Science by examining the key skills sought by employers and analyzing salary trends across various regions of Australia. The intended audience could be Data science students and current professionals looking to transition into data science roles. It aims to provide insights into the job landscape, including the skills sought by employers and salary trends. As data-driven decision-making becomes increasingly prevalent across industries, understanding the dynamics of data science jobs is essential for both aspiring data science students and current professionals to make informed decisions regarding their career paths and skill development strategies.

Data Sources

The data source for exploring Data Science job opportunities comes from Kaggle. The data was collected, prepared and cleaned by Nomilk from job listings sourced from seek.com.au, a prominent Australian job search website. It provides over 50 columns of rich numeric, geographic and text data to explore.

This dataset contains information related to job listings, including various qualitative and quantitative variables.

Qualitative Variables: These variables describe qualities or characteristics and include job title, classification, advertiser name, company name, teaser, nation, state, city, area, suburb, work type, whether right to work is required, desktop and mobile ad templates, company profile URL, seek job listing URL, first seen, last seen, and recruiter.

Quantitative Variables: The dataset also includes quantitative variables such as listing date, expiry date, salary string and skills.

Project Plan

The scope of this project is to provide insights into the Australian data science job market for both data science students and current professionals. This will be achieved by visualizing key attributes from a dataset extracted from job listings that includes analyzing distribution of jobs across different states, identifying key skills employers are seeking, examining salary trend over the time, trend in job demand and availability over time to provide insights into the dynamic nature of the job market. These insights could be useful for making informed decisions regarding the career path and skill development strategies.

Elementary Question

- Which state has most job listings per 100k population across different states in Australia?

Intermediate Question

- What are the top three programming languages in high demand in Australian data science job listings?
- Which states have the lowest and highest salary for Data Science jobs in Australia?

Overall Question

- What is the trend over time regarding the demand and availability of jobs in the Australian data science market?
- How does the average salary for Data Science jobs in Australia varies over time?

Handling missing data poses one possible problem, as incomplete information could affect the accuracy of the analysis. Additionally, joining the dataset with other sources may present challenges, particularly in ensuring data compatibility and consistency between sources. These tasks require careful attention to detail and thorough data validation to maintain the integrity and reliability of the analysis results. Selecting the right visualization tool and learning its features is crucial and may be time-consuming. Therefore, it should be dealt with care.

Data Preparation

To enhance the analysis, the job listings data is augmented with state population data obtained from Wikipedia. This additional information facilitates a deeper understanding of job opportunities relative to the population density of each Australian state and territory. The process involved standardizing state names, aggregating job listings by state, and normalizing them with state populations retrieved from the source. This combined dataset allows for insights into the distribution and density of data science job listings across different regions of Australia. Additionally, date modifications were done, including converting listing and expiry dates to date format and creating new variables for month and year.

Data Exploration

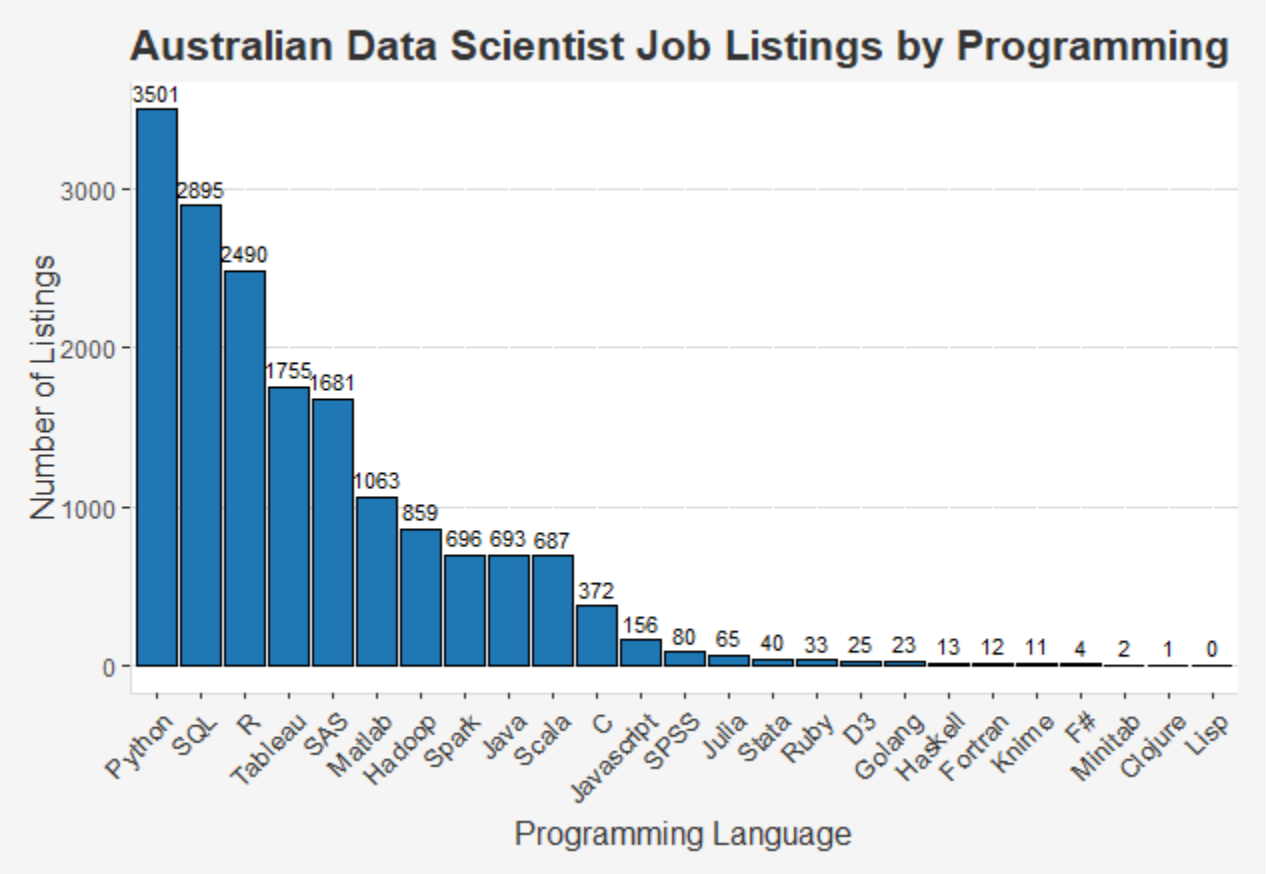


Figure 1: Australian Data Scientist Job Listings by Programming

- The top three programming languages in high demand in Australian data science job listings are Python, SQL, and R.
- Python emerges as the most sought-after language with a total of 3501 job listings. It is followed by SQL with 2895 listings and R with 2490 listings.
- These findings highlight the essential role these languages play in various aspects of data science projects and their high demand suggests that proficiency in these languages is crucial for success in the field of data science in Australia.
- These preliminary results provide valuable insights into the top three programming languages in high demand in Australian data science job listings. (Intermediate question)

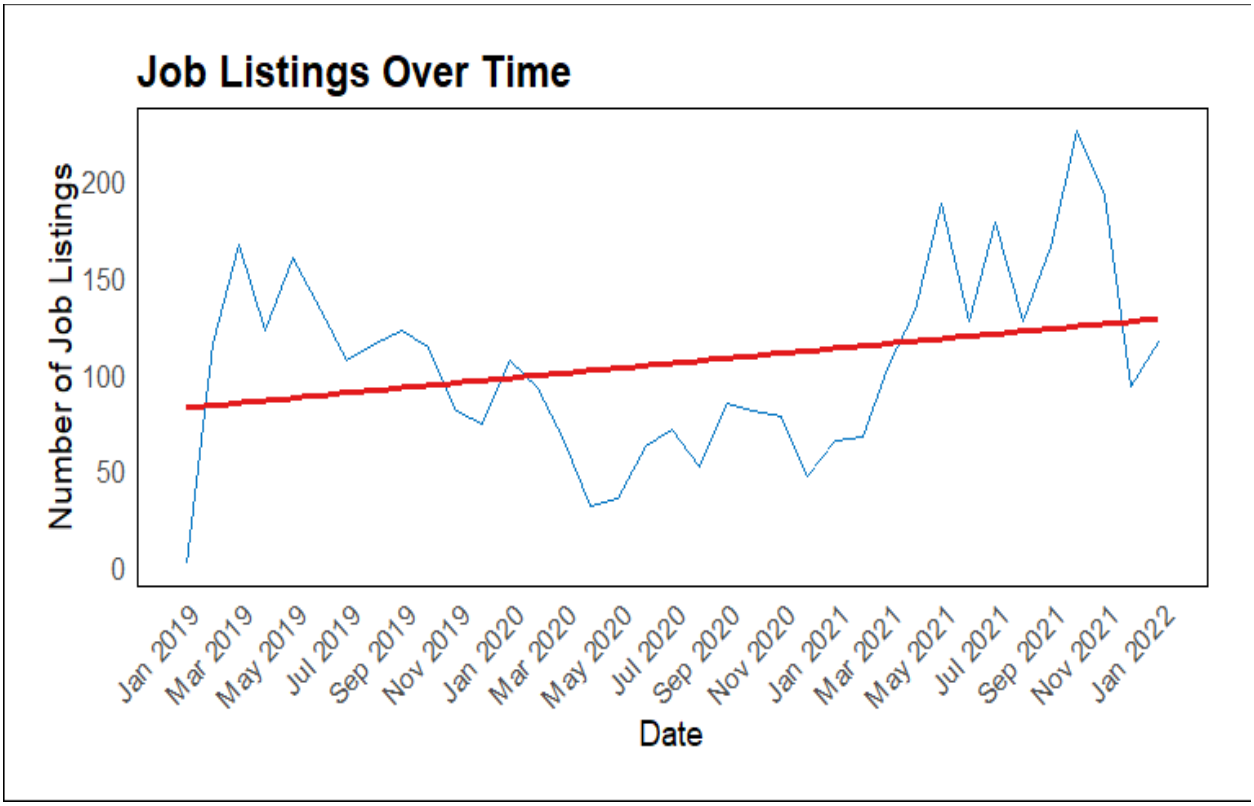


Figure 2: Job Listings Over Time

- It is evident from Figure 2, that the demand for data science jobs in Australia has been steadily increasing over time.
- The trend line indicates a consistent upward trend in job listings, suggesting a growing demand for data science professionals across various industries.
- The increasing number of job listings reflects the expanding opportunities in the field of data science and underscores the importance of data-driven decision-making in today's business landscape.
- These preliminary results provide valuable insights into the trend over time regarding the demand and availability of jobs in the Australian data science market. (Overall question)

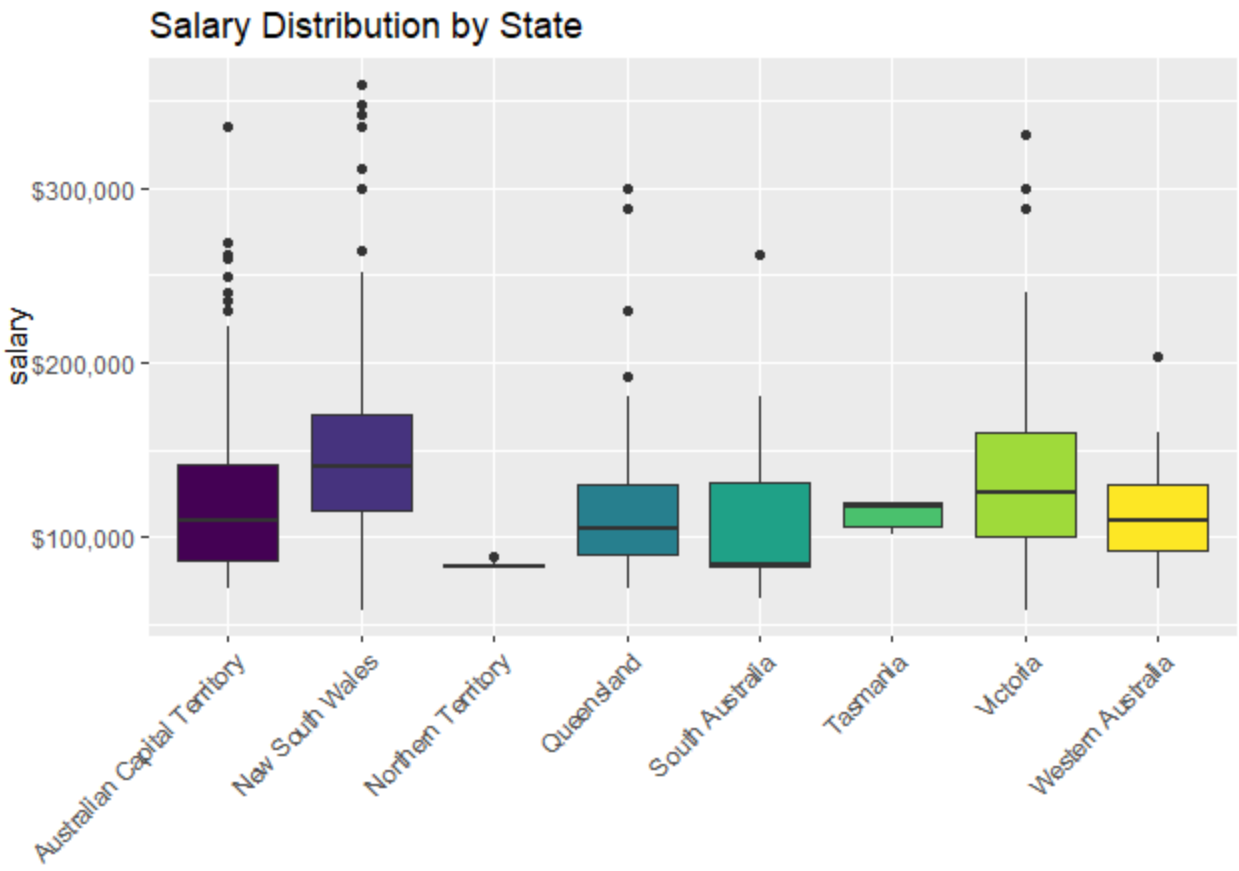
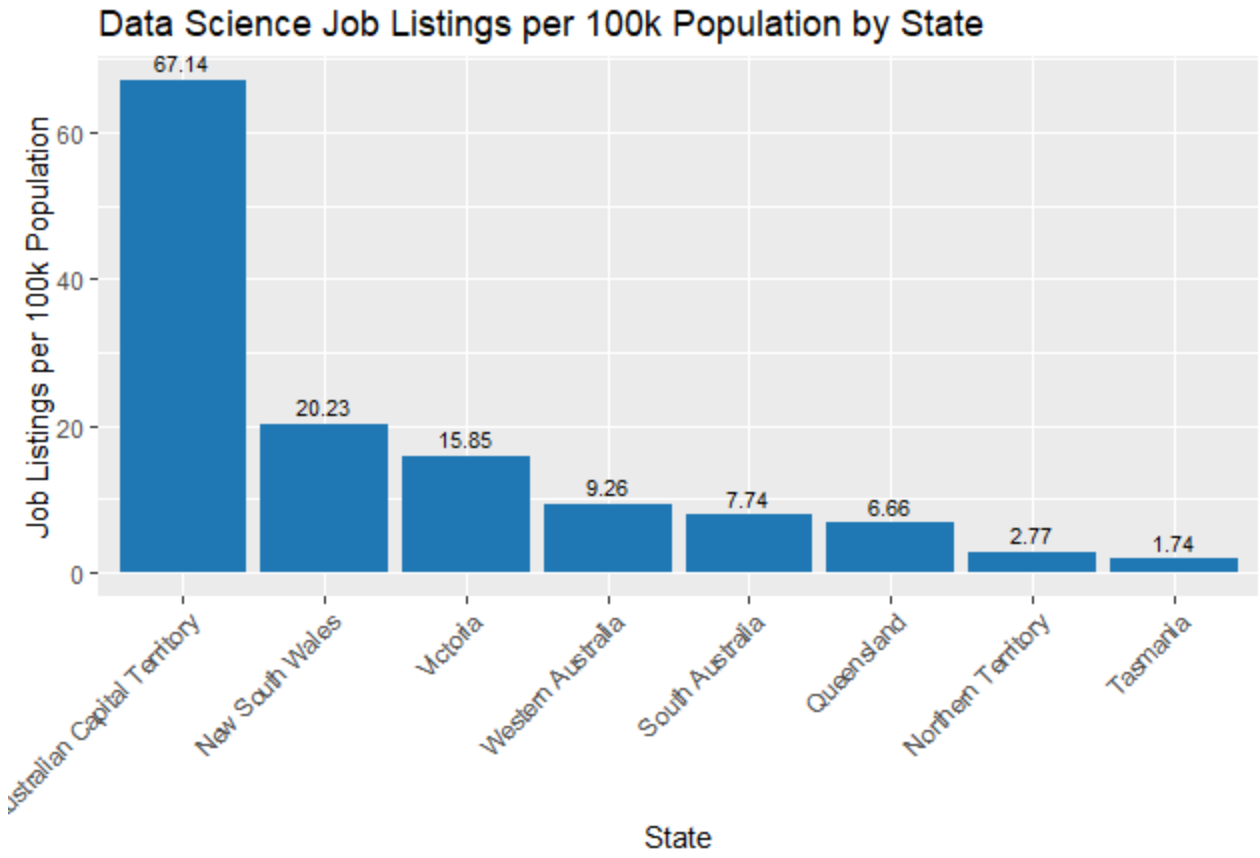
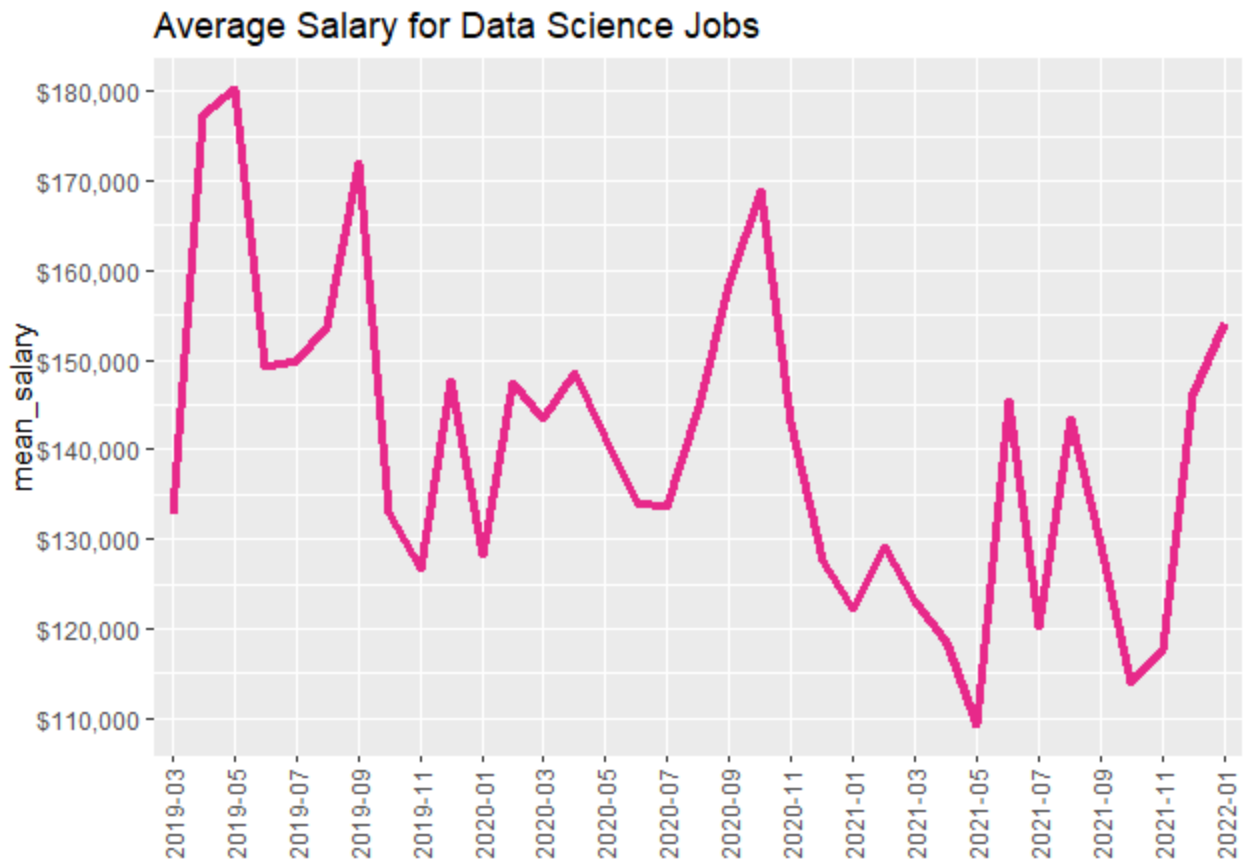


Figure 3: Salary Distribution by State

- From Figure 3, it's apparent that Victoria has the highest variability in salary for data science jobs, surpassing that of other states. It is followed by New Soth Wales. Conversely, the Northern Territory exhibits less variability in salary compared to other states. Outliers are observed across all states, indicating instances of exceptionally high or low salaries.
- Based on this analysis, Victoria appears to have the highest average salary with the highest variability, suggesting a wide range of salary offerings for data science roles in the state while the Northern Territory has a comparatively lower average salary with less variability.
- These preliminary results offer valuable insights into which states in Australia have the lowest and highest salaries for Data Science jobs. (Intermediate Question)



- According to Figure 4, the Australian Capital Territory (ACT) boasts the highest number of job listings per 100k population, with a percentage of 67.14%.
- It is followed by New South Wales with 20.23% and Victoria with 15.85%.
- These findings suggest that the ACT has a significantly higher concentration of data science job opportunities relative to its population compared to other states in Australia and also provide valuable insights into the distribution of data science job opportunities across different states in Australia.
- These preliminary results offer valuable insights into which state has most job listings per 100k population across different states in Australia. (Elementary Question)



The trend in the average salary for Data Science jobs in Australia fluctuates over time, as indicated by the data. From March 2019 to July 2019, the average salary shows a rising trend. However, there is a dip observed until July 2020, followed by a rise until November 2020. Another dip is noted until March 2021, after which the average salary starts rising again. These fluctuations suggest that the salary dynamics in the Australian Data Science job market are subject to various factors and trends over time. These preliminary results provide valuable insights into how the average salary for Data Science jobs in Australia varies over time. (Overall question)

References

1. Ramasubbareddy, S., K, G., Luhach, A. Kr., & E, S. (2019). Big data analysis on Job Trends using R. *Recent Patents on Computer Science*, *14*(1). https://doi.org/10.2174/2213275912666190801113359
2. Data Science Job Listings - Australia - 2019-2022, Kaggle <https://www.kaggle.com/datasets/nomilk/data-science-job-listings-australia-20192020/data?select=listings2019_2022.csv>
3. State population Data <<https://en.wikipedia.org/wiki/States_and_territories_of_Australia>