3803ICT Big Data Analysis Assignment

Part 3 - Evaluation

Joshua Russell | s5057545
joshua.russell2@griffithuni.edu.au

Joshua Mitchell | s5055278
joshua.mitchell4@griffithuni.edu.au

1) What are the findings of your data analytics?

Our main two sections of analysis involved investigating the market by location, and then by sector. While studying the market by location, we found that Sydney has the greatest job market share among all cities of Australia, making up 30.8% of the total job listings on SEEK. Furthermore, the combined market share of Sydney, Melbourne and Brisbane was found to be 65.4%, highlighting that these three cities offer the majority of jobs in Australia. Refer to *Figure 1* for one of our data visualisations of this study.

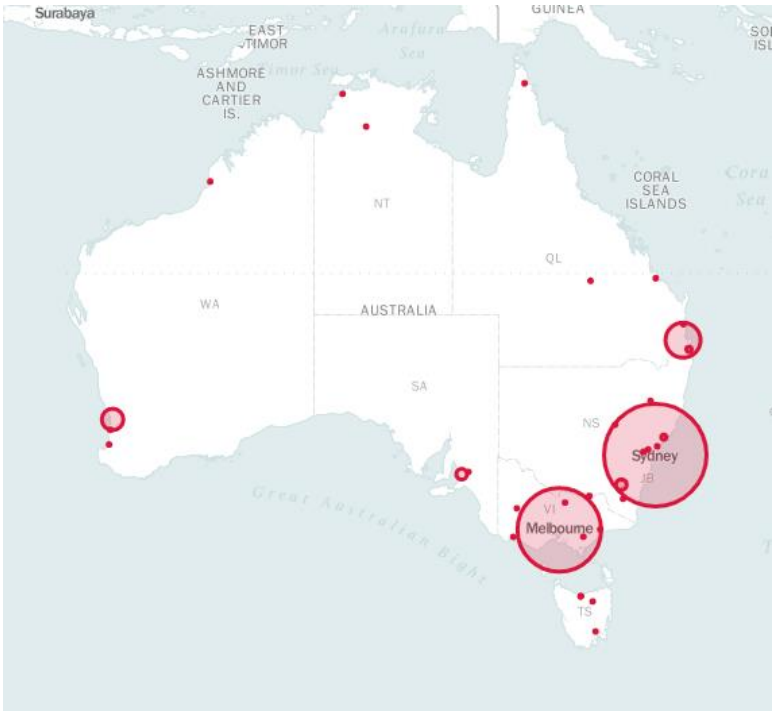

Figure 1. Map representation of the Relative Market Size for each city in Australia

We then investigated popular sectors in each city and found that Information Communication Technology is the hottest job sector in the major cities of Australia (i.e. Sydney, Brisbane, Melbourne and Canberra). In comparison, among the other non-major cities the most popular sector was Healthcare/Medical. Further observations were made, such as that the Trade Services sector was frequently within the top 3 most popular sectors of each city, although rarely the most popular, and that cities had popular sectors depending on their attraction (such as Gold Coast with Hospitality & Tourism) or on their commodity (such as Perth with Mining Resources & Energy). Studying the hottest sectors within each city was also beneficial for understanding the distribution of job sectors across Australia. It was observed that the Information Communication Technology sector had the largest market share

across Australia. However, by effectively observing the market share in each city, we found that the majority of cities in Australia have Healthcare/Medical as their most popular job advertisement sector. This observation alludes to the fact that the Information Communication Technology sector is predominant within only a few cities, where the vast majority of job offerings for the sector are within those cities and not across Australia. A select number of pie charts are presented in *Figure 2* as a summary of this investigation.

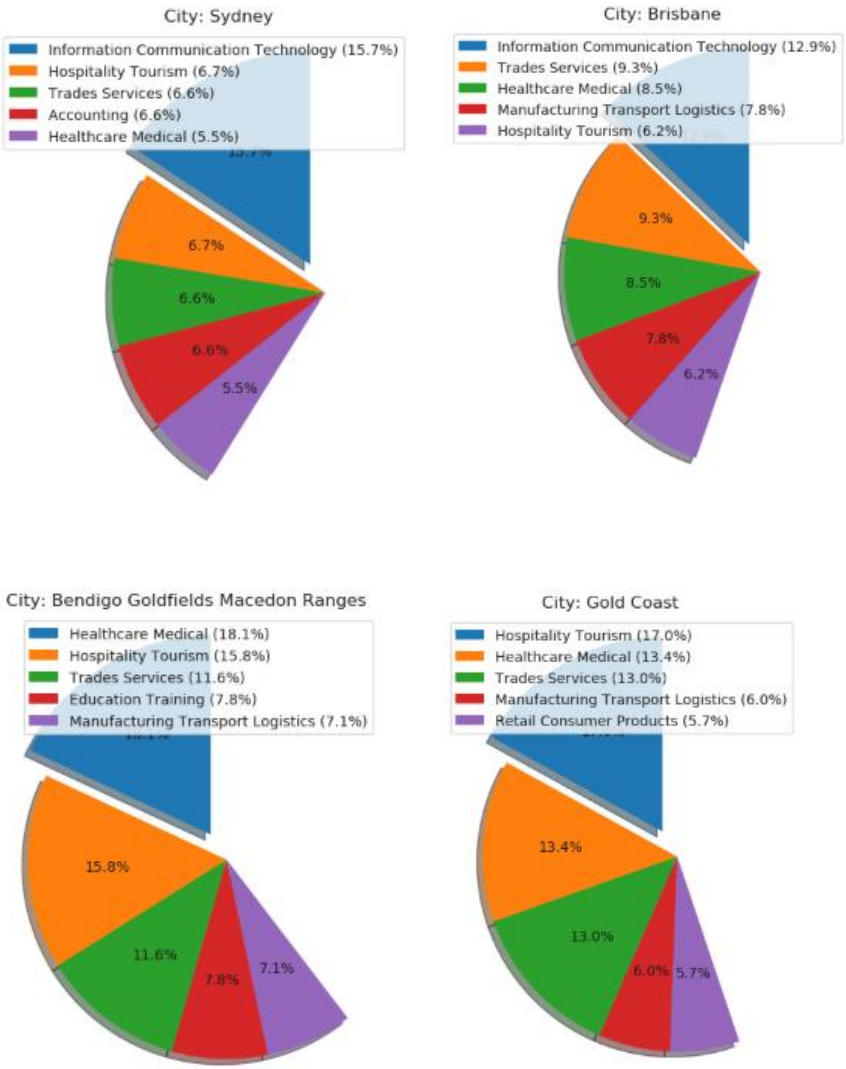

Figure 2. Percentage Market Share of the Top 5 Sectors within Cities of Australia

Our following investigation examined where in Australia employees are more well paid. The metric we used for assessing this was the average salary per location. We found that the ACT is the most well-paid location in Australia, having a widespread distribution of salaries within the location. The ACT does not have near the largest population in Australia. However, in comparison to the other locations, its salary distribution is skewed towards the higher end of the salary range. Since the ACT is where Australia's capital is, and where the Parliament House is, it is assumed that this higher distribution of salaries is due to governmental jobs generally paying higher. Interestingly, the location with the second highest average salary was found to be Port Macquarie, followed by Sydney. *Figure 3* presents the top 10 most well-paid locations in Australia according to the metric of average salary per location.

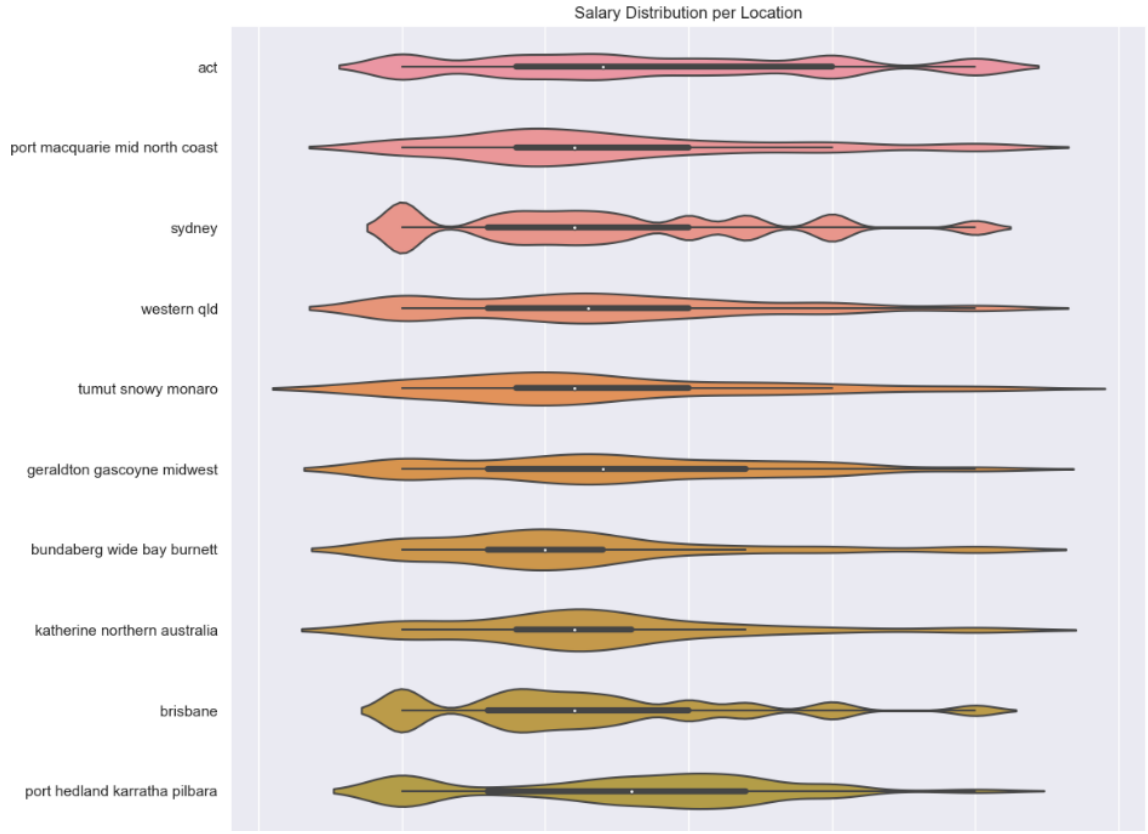

Figure 3. Violin Plots showing the Salary Distributions of the Top 10 most Well-paid Locations in Australia

The next investigation looked at the trend of salary over the 6 months that the data was collected. Our results showed a slight decline in overall job salary over the 6 months. The largest peak was around December of 2018 with an average salary of $160,000+, followed by periods of low peaks of $0 - $30,000 corresponding with relatively low standard deviation. This is an interesting finding and may be due to companies hiring large amounts of people for the new year and over the Christmas holidays. Generally, we also see large spikes on certain days that could be the result of large companies releasing new positions at the same time. For example, Coles hiring people periodically across multiple locations. Later when looking at which companies post most frequently, we saw other job search websites like Jora Local post their own listings on SEEK. Consequently, this could also be the cause for the large spikes. To improve our results and gain further insight we would need to collect data over a longer period. Collecting over a longer period will improve reliability and reduce the impact of noise on our results.

Our following study was to detect the pattern of job posting. The pattern of job posting is an important source of information to uncover and understand as finding a job often is partly being in the right place at the right time to submit an application. Knowing when jobs are generally posted can be helpful to know when to prepare sooner and find jobs on time. We found, as hypothesized, that weekdays have the greatest number of job posts, averaging 61,182.6, with Wednesday as the most common overall. The average of weekdays and weekend days differs by 50,651.6. Furthermore, weekdays differ at most by 14,663, which was between Monday and Wednesday. Interestingly, we saw a decline in a staccato like fashion in the number of job posts from October to March. Months December and January show an uptick of around 30,000 compared to November, likely because of the Christmas and holiday hiring period. Even more interesting is the decline in job postings in March down to 402 and large increase in October to 113,782. October may be the Month when most companies post job listings so that they can

have time to prepare for people to start working the following year. March in contrast is after when they have already hired for the year. We cannot be completely certain as we do not have a year's worth of data to draw the whole seasonal period from to determine if this is an anomaly or not. We also found a pattern of jobs being posted within the first or second week of each month. The only exception being October where the largest peak is in the 3rd week. *Figure 4* presents the graphical results of this analysis for the days of the week.

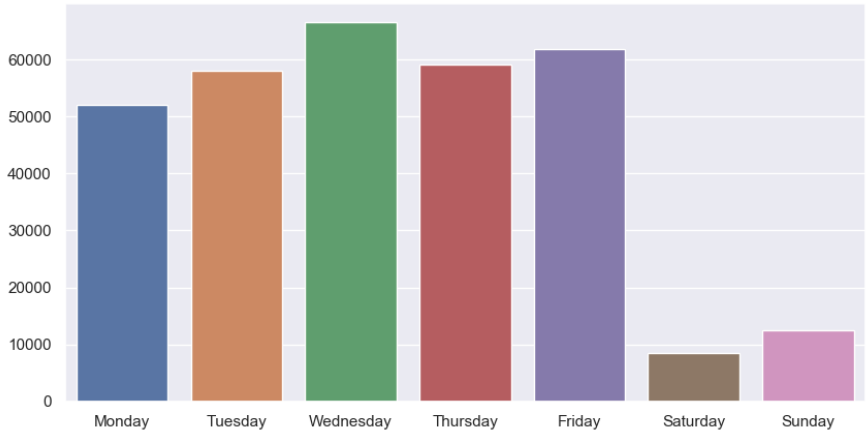

Figure 4. Frequency of Job Postings for each Day of the Week throughout the 6 Month Period

The second main section of the analysis looked at studying the market by sector. We initially found that the sector with the largest market share, and thus the trending sector in the market, was Information Communication Technology. Followed by Trades Services and then Healthcare/Medical. The sectors with the lowest market share were Self Employment, Advertising/Arts/Media, and CEO/General Management. Which suggest that the sectors are either not in demand (assumedly Advertising/Arts/Media), or job offerings within that sector are rare (assumedly CEO/General Management). *Figure 5* summarises the findings of this analysis.

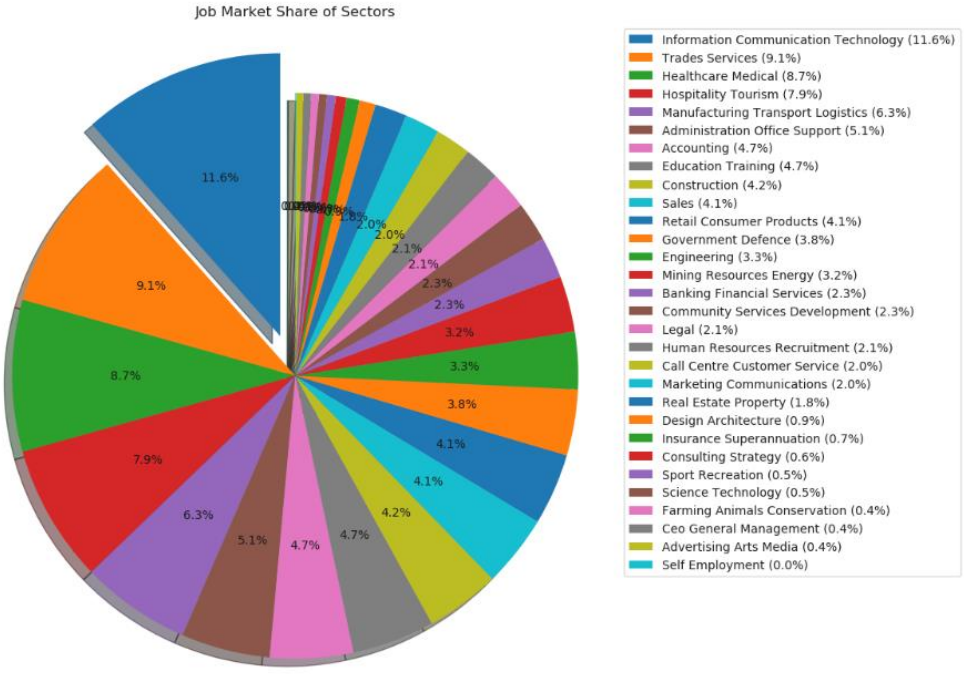

Figure 5. Percentage Job Market Share for each Sector

The subsequent analysis further examined the market share by looking at which sub-sectors within each sector are the most popular. This investigation allowed us to determine the most popular sub-sectors within each sector. For example, Automotive Trades were the most popular in Trades Services, and Nursing Aged Care was the hottest sub-subsector within the Healthcare/Medical sector. In addition to this, we were able to examine how much of the sector market share was taken up by the top 5 sub-sectors within each sector. For example, in the Retail Consumer Products sector, the top 5 most popular sub-sectors made up 94.3% of the total sector market share. Whereas in the Healthcare Medical sector for example, the top 5 most popular sub-sectors contributed to only 43.4% of the total sector market share. We consequently concluded that in sectors such as the Retail Consumer Products sector, there are limited options within the sector in regard to diversity of sub-sector. In comparison, for sectors who's top 5 sub-sectors made up a rather small portion of the total market share, such as in the Healthcare/Medical sector, we concluded that there exists many more subfields with decent market shares and therefore more opportunities for diversity within those sectors. *Figure 6* shows this comparison between sectors and how much their market share is dominated by their top 5 sub-sectors.

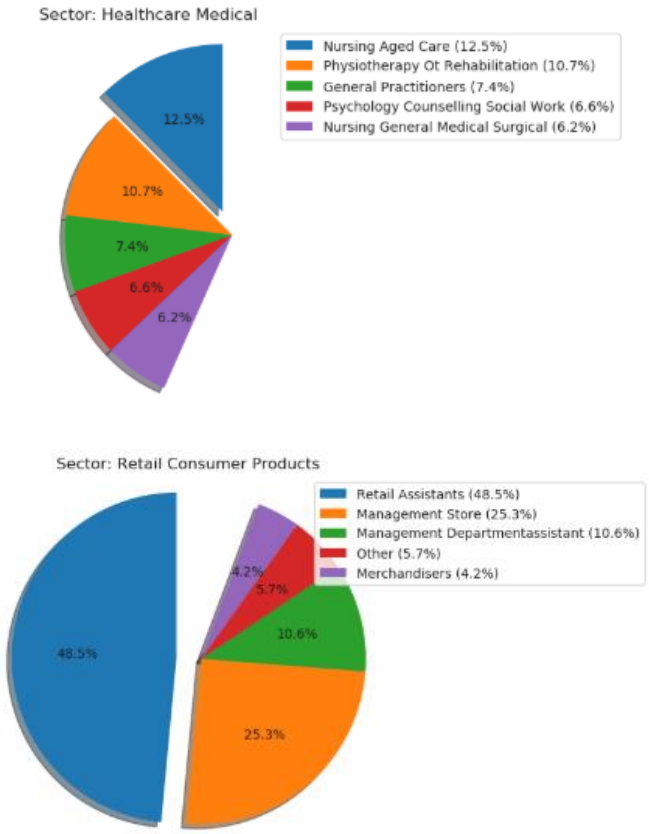

Figure 6. Percentage Market Share of the Top 5 Sub-sectors within two Sectors

We then investigated the salary range for each sector and sub-sector. We found that all sectors had an annual salary range of $0 to $200,000+. However, when looking at the annual salary range for sub-sectors, we observed some differences. We found sub-sectors with small salary ranges, such as Housekeeping with an annual salary range of $0 to $80,000, and Nannies/Babysitters with the smallest sub-sector annual salary range of $0 to $60,000. In comparison, sub-sectors such as Oil Gas Drilling, Aerospace Engineering and Intellectual Property Law all had salary ranges of $0 to $200,000+. Our conclusion in regard to this was that the level of expertise required to work in the sub-sector is a more influential factor for determining the salary range of a sub-sector than the demand for the sub-sector in the current job market.

To compare these annual salary ranges between sectors, we conducted another investigation observing the distribution of sub-sector salary ranges within each sector. This showed that the CEO/General Management sector had the majority of its sub-sector job listings within the annual salary range of $120,000 to $200,000+. In comparison, the Retail Consumer Products and Hospitality/Tourism sectors had significantly more job listings than the CEO/General Management sector, however the majority of the listings were in the lower salary ranges. We also observed that the Information Communication Technology sector's sub-sector job listings were skewed towards the salary range of $200,000 to $200,000+, highlighting that a significant number of the job listings within the sector were well paid. The complete graphical representation of this study is presented in *Figure 7*.

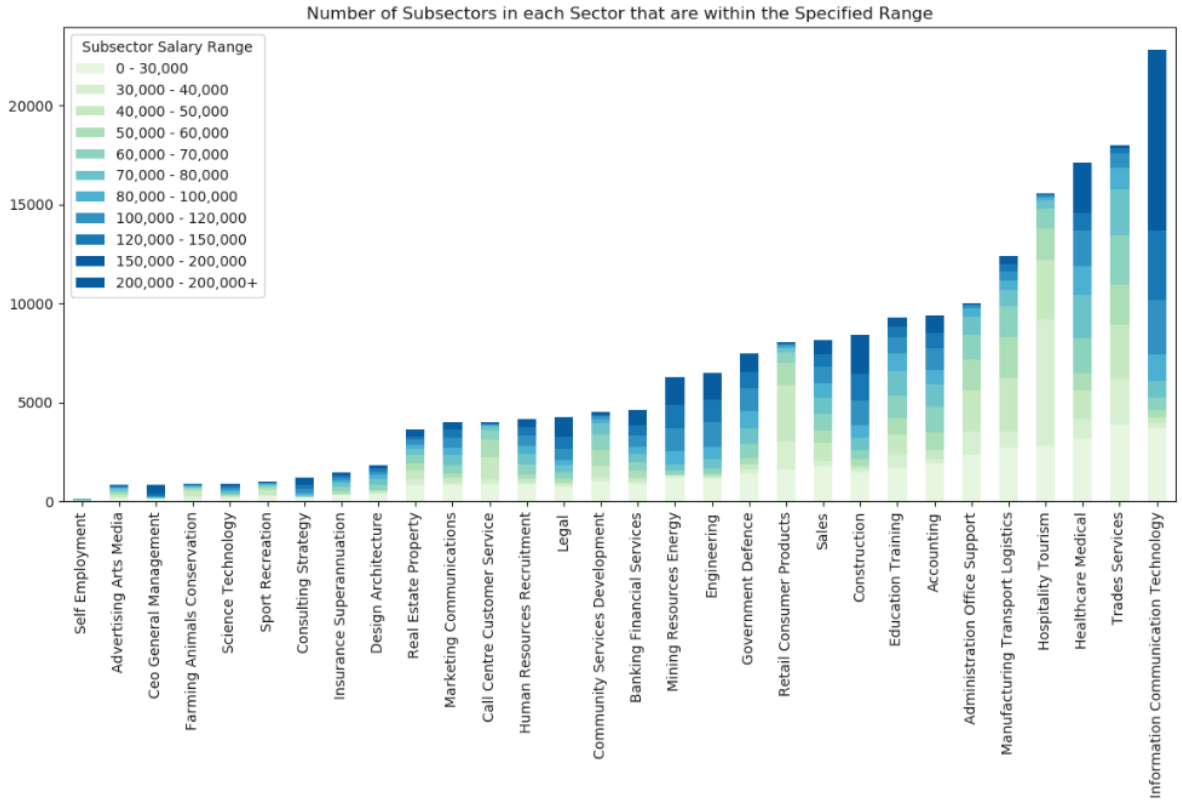

Figure 7. Number of Sub-sector Job Listings in each Sector that are within a specified Annual Salary Range

Following this investigation, we analysed what skillsets are required and/or desired within each sector. We were able to successfully extract relevant skillsets for all of the sectors within the dataset. For example, within the Education Training sector we found that the top skill required is the ability to work with children. Furthermore, for the Sport Recreation sector, the most frequently requested skillset in job listings was first aid - CPR. We then examined the top skillsets among all sectors combined and found that the most popular skillset among job advertisements on SEEK was communication skills, followed by customer service. Interestingly, having a driver's license (or in regard to skillsets, having the ability to drive), was ranked fourth, above skills of management, being a team player and having time management skills, which were all popular among the job listings investigated. The skillsets within each sector were then compared, and we were able to find that sectors which seemed to be very different in practice still had a lot of desired skillsets in common. A graphical representation of this comparative analysis is presented in *Figure 8*.

Figure 8. Scatter Text Graph comparing the Skillsets of the Trades Services and Information Communication Technology Job Sectors

2) What actions for balancing the markets do you suggest based on your findings?

Significant disparity in the job market was mainly found in the location of job offerings, as a few major cities in Australia held the majority of the market share. It is hypothesised that this is the case due to companies having offices within the CBDs of these cities and therefore are offering the majority of jobs. However, the population size of these cities would also contribute to this factor, as a larger population within an area naturally causes the number of job offerings to rise. In order to balance the market based upon this finding. the government could support local businesses more, such that they develop and start offering more jobs. This way people will not need to move to the major cities of the country to find worthwhile work. Moreover, the government could invest more into developing the CBDs of the other main cities in Australia, such as Perth, Darwin, Adelaide, and Gold Coast for example, so that more people move to those locations to live and companies gain an interest in investing in those areas. If companies began to build more offices in those areas to further develop them, then the market share in regard to location should become more balanced over time. Although there will almost always be some disparity due to some cities being more develop than others.

The other main point of disparity in the job market was in regard to job sectors. With sectors such as Information Communication Technology having 11.6% of the market share, and other sectors such as Sport/Recreation with only 0.5% of the market share. It is assumed that this is a result of current trends in the marketplace, as humanity is going through the Information Age. Moreover, as technology becomes more advanced and artificial intelligence begins replacing people in their jobs, it is hypothesised that this disparity will become more pronounced. The government therefore needs to offer move incentives to keep the market balanced such that people can find and work in jobs that give them fulfilment. The government can invest in sectors on a case by case basis to assist in balancing out this disparity in the market. For the Real Estate/Property sector, government incentives can be put in place to encourage people to buy and sell more. The government can invest in developing rural areas, proving more jobs to Trades Services workers and then in turn Real Estate agents. With more properties on the market, real estate agencies will start offering more jobs, which will therefore assist in balancing the market share among sectors. Although, with this approach there will almost always be disparity in the

job market as we can observe from history. We therefore hypothesise another approach, that if companies and the government invest in Information Communication Technology for a few decades, after artificial intelligence has advanced to a state of general intelligence, people will no longer have to do a significant number of the "standard" jobs we have today. This may lead to people being self-employed or going into more personal endeavours. With everyone doing their own thing, based on their own interest, we may see the greatest balance in the job market we have ever seen. Thus, this option is another potential action that can be taken to balance the marketplace in the future.

3) How could you refine your data analytics?

As in any data analytics investigation, our data analytics could be refined by using more data in our investigations. Since we are effectively using a sample distribution for our analysis, any additional data will assist in approximating the underlying population distribution of the study, consequently improving the accuracy and validity of the conclusions reached and making outliers and noise easier to detect. The first step in collecting more data would be to investigate and use different data sources. Since this investigation is on a dataset from the employment marketplace domain, additional employment marketplace websites could be used for gathering data (such as Indeed, Jora, and CareerOne for example). Not only could additional data sources be used, but the existing sources could be further analysed for more data. The SEEK dataset used within this study contains the information of 318,477 job listings, recorded from October 2018 to March 2019. As a result, the SEEK website could be crawled for additional data from the most recent months.

Increasing the dataset size for the study will assist in the accuracy and validity of the investigation, however it will not provide additional data attributes to investigate. The downside to the SEEK dataset used within this study was that the only meaningful numerical attributes were the lower and upper salary ranges of the job advertisements. Moreover, since these attributes only represented the salary range for a particular sub-sector, there was no unique salary for each of the jobs. This consequently limited the depth of analysis when examining the annual salaries of job listings. The SEEK website offers a "career insights" tab on their job advertisement pages which states the projected job growth over the next 5 years for the sub-sector. This information would provide interesting insights into the sub-sectors and provide another method for our analysis to be refined. Furthermore, numerical data in relation to the annual earnings of the companies advertising the job listings, as well as simply the salaries of the job listings would provide numerical data to perform statistical analysis methods on, that would in turn refine our investigation. Gathering this information would involve tweaking the web crawler used to gather the original dataset and using other sources to find details about the companies posting such information.

Another valuable source of information would be the resumes of the applicants to each of the jobs, with information on which of these applicants ended up getting the job. This type of information would assist in refining our skillset analysis of the job sectors, since we could analyse the characteristics and achievements of the applicants in order to see what were the necessary skills that lead to them being hired. This information would be much more difficult to obtain. However, it would be very valuable for improving the results of our study and investigating additional natural language elements of job advertisements.

4) Are there any implications for employers and employees based on the findings you obtained?

The first implication that is evident in our findings is that of salaries within locations. Employers operating businesses in locations that take up a significant portion of the market share, say for example Sydney with 30.8% of the job listing market share, will be in a more competitive environment when seeking employees. If the employer's business is in the Information Communication Technology sector, then they will need to offer higher salaries to secure employees than if they were in a location that did not contribute so much to the total market share. This also means for employees that if they are seeking a job in a popular location, they will have higher salary options. However, they will also be competing against more people for the job.

Our findings in regard to the which sub-sectors are the most popular within each sector suggested an interesting fact on sector diversity. We found that within some sectors, the top 5 sub-sectors contributed to over 90% of the total sector market share, whereas in other sectors, the top 5 sub-sectors contributed to less than 50% of the total sector market share making them more diverse. This implies that employees working within these diverse sectors have more options in regard to their future career path within that sector. Meaning that if they want a change, they do not have to move into a whole other sector of work. These diverse sectors may also be favourable for employers to get into, since they can recruit employees from other sub-sectors and therefore have more options for employment. Examples of these diverse sectors from the investigation are Accounting, Trades Services, Healthcare/Medical, and Banking/Financial Services.

The investigation on the number of sub-sectors within a specific salary range for each sector showed the distribution of potential salary ranges over a particular sector. This has its implications for employees, as if they are in a lower salary range job within a sector that has a distribution of jobs within all of the investigated ranges (i.e. from $0 - $30,000 up to $200,000 - $200,000+), then the employee has the opportunity to move up the salary ranges in their career, earning a better living for themselves as they improve in their profession. Examples of these sectors are Information Communication Technology and Healthcare/Medical. In contrast, employees working within sectors that were found to mainly just have lower salary range positions can find the limit in salary of their career and make choices from there depending on their interests and ambitions.

Lastly, the study in regard to the top skillsets of sectors and of the job market as a whole has significant implications for employees. Employees are able to examine the most frequently requested skillsets within their sector and work towards perfecting them or obtaining them if they are yet to gain that skill. This offers employees with a path of progression and development they can work towards in their career, knowing that the traits they are acquiring are worthwhile within their field of work. In addition to the skills of their sector, they can also observe the top skillsets among all sectors. These skillsets somewhat reflect what is required, and also what is desired in an ideal employee for most employers. Traits such as strong communication skills, being a team player, and knowing how to manage projects. These can be taken on by employees in any sector to assist them in their career development, as they are general skills that are clearly wanted by employers everywhere.

5) Present and visualize your data story on an online webpage

We created a webpage for presenting our data story. It contains visualisations that are both static and interactive to highlight, summarise and explain our analysis of the SEEK dataset. The interactive data visualisations were implemented to allow viewers to explore the parts of the analysis they are interested in, and to make the experience more enjoyable and understandable. The website was published on Google Cloud, and can therefore be accessed through the following link:

https://stage-dot-linuxvm-216704.appspot.com/