Which Programming Language will be more in demand for future jobs

An analysis of software developer jobs, its wages , importance of non-technical skills

Shubham Gupta
Department of Computer Science
University of Adelaide
Adelaide, SA, Australia
Shubham.gupta@student.adelaide.edu.au

ABSTRACT

According to Wikipedia, there are about 700+ programming languages out there[1] with varying features. Learning a programming language takes time and energy and where you should invest your time and energy is an important question you should be asking yourself. In this paper I tried to answer this question and few more with the help of data. All of the data collected are from jobs posted within Australia on websites like Seek, LinkedIn, Indeed. We also check how the job pays and try to relate a respective programming skill to decide if there is a programming language that pays better than others, you don't want to learn a programming language which is not going to return you investment. Finally we will be checking if technical skills is all you need to qualify for job, we will be analysing the data from the employers in what they seek in potential employees. All this will hopefully help you decide the skills you should learn and focus on.

1 INTRODUCTION

According to Wikipedia, there are about 700+ programming languages out there[1] with varying features. Learning a programming language takes time and energy and where you should invest your time and energy is an important question you should be asking yourself. And in these uncertain times investing in the right thing might just stand you apart from the rest. "The Competitive pressures, globalization, and changing technology are causing firms to re-evaluate the process of how work is done" [2]. Carnevale et al.[3] talks about "shift from the industrial era to the post-industrial era of the knowledge economy" and how this resulted new skill requirements from companies, technical skills augmented with growing need for non-technical skills like "learning, reasoning, communicating, general problem-solving skills and behavioral skills". James J. Cappel (2002) [9] conducted a survey for his research which included employers that had recruited students from their university in last three years.

2 MOTIVATION

As a university student graduating in about a year, the questions like what skills do I need to get a good job always linger my mind and I am confident enough to say that I am not alone. You can find bits and pieces of information on the various articles but the credibility of their data is always questionable and does this data apply to you, is also unclear. Also we thousands of options to choose from when we want learn a new skill, be it a particular course at the University or any online course. We always think if this particular skill will help us in the future or not. Most people specially in Software Engineering never focus on non-technical skills, examples include soft skills. RQ3 will answer if non-technical skills are important or not. Do employers look for them in individuals or not. In conclusion, all these research questions posed are important as it will help other people (specially students) in software engineering understand which skills are important for them and what employers are looking for.

3 BACKGROUND

3.1 Seek

The main website we will be using to scrape the data of software development job available in Australia will be Seek.com, "Seek Limited and its subsidiary companies, known as the Seek Group, focus on facilitating the matching between jobseekers and employment opportunities and helping hirers find candidates for advertised roles"[4]. "SEEK makes a positive impact on a global scale, with exposure to 2.9 billion people, more than 51 million students and learners and a presence in 18 countries including China and across South-East Asia and Latin America"[5]. Because of it's impact on the students and website being originally conceived in Australia, and many of local Australian companies directly posting jobs on it, this is a perfect choice to scrape the data from and get a good idea of the local landscape.

3.4 Software Used for Analysis

3.4.1 Anaconda Navigator

Anaconda is a free and open-source distribution of the Python and R programming languages for scientific computing (data science, machine learning applications, large-scale data processing, predictive analytics, etc.), that aims to simplify package management and deployment. The distribution includes data-science packages suitable for Windows, Linux, and macOS[6].

3.4.2 Jupyter Notebook

Jupyter Notebook (formerly IPython Notebooks) is a web-based interactive computational environment for creating Jupyter notebook documents. The "notebook" term can colloquially make reference to many different entities, mainly the Jupyter web application, Jupyter Python web server, or Jupyter document format depending on context. A Jupyter Notebook document is a JSON document, following a versioned schema, and containing an ordered list of input/output cells which can contain code, text (using Markdown), mathematics, plots and rich media, usually ending with the ".ipynb" extension[7].

3.4.3 Packages – Beautiful Soup

Beautiful Soup is a Python library for pulling data out of HTML and XML files. It works with your favorite parser to provide idiomatic ways of navigating, searching, and modifying the parse tree. It commonly saves programmers hours or days of work[8].

We will use the beautiful soup package directly in our Jupyter notebooks, which involves calling and parsing our desired website and then extracting the data. One of the major advantages of using Jupyter Notebooks in Anaconda is the ease of use. We can extract and then analyze the extracted data in one single file. At the end the paper, I will the sharing the respective .ipynb file on GitHub so that data extracted can be recreated and verified, and someday even be used to improve upon the current model.

4 RESEARCH METHOD

The main sources of data required to answer RQ1 and RQ2 are job postings on certified websites like LinkedIn, Seek and Indeed. For RQ3 we will be using survey conducted by James J. Cappel (2002) [9], which had answers from different organizations, some local and some global about the quality that the employers seek in individuals.

4.1 Research Questions

There are three main questions I want to answer from this research paper, firstly *"What will be the top skill requirement in the near future, in terms of the programming skills"*. For this question we will be mainly focusing on the programming skills requirement in Australia, otherwise the data collection for this question will be endless. But if we narrow our data collection to one country, it will give us more cohesive results and will be helpful to the local community, specially in this case the current student who will be benefitted the most with the findings of research paper.

The second question is *"How are these skills related to the wages?"*. This question can be subjective, i.e. may change depending on individual but we are going to answer this question in terms of how much you are going to earn from that skill, from jobs you can get with those skills. It will give some quantifiable parameters to individuals reading it, help decide if they want to learn a specific skill or not.

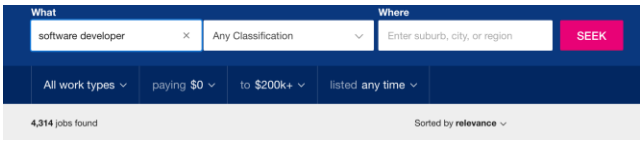
The third and the final question is *"Importance of non-technical skills compared to technical skills"*. Most people specially in Software Engineering never focus on non-technical skills, examples include soft skills. This question will answer if non-technical skills are important or not. Do employers look for them in individuals or not.

In conclusion, all these research questions posed are important as it will help other people (specially students) in software engineering understand which skills are important for them and what employers are looking for.

4.2 Data Collection

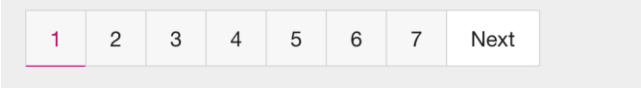
4.2.1 For RQ1 & RQ2

What we wanted to collect as our data is the entire job description from websites such Seek.com. These website have a search option generally on the home page where you can type the profile or field you are looking for as shown in Fig. 1. If you enter the following keyword (in our case we are looking for software development jobs in Australia) we get 4314 jobs results[10] as shown in figure below.



As you can imagine going through all the jobs one by one will be tedious task and will be very hard to keep track of, so therefore we are going to write a python script for it. As mentioned in section 3, We will be using Jupyter Notebooks in Anaconda Navigator for this as this gives the right balance between running a program and analyzing data we have extracted.

Now we will start writing our code to get data from all 4314 jobs which is basically going inside every job posted (4314) by employer and extracting the description of the job.
Our first problem was that not all jobs were posted under one link, what I meant by that is all of the 4314 jobs are divided between multiple pages as shown below.



The link for this page one is “https://www.seek.com.au/software-developer-jobs”. If we move to page 2, the link becomes “https://www.seek.com.au/software-developer-jobs?page=2”. If we continue the link follows a pattern which we can automate using a program, so we will start with that, writing a program which gives us all the links on which our main extraction program will run. Also there are about 20 jobs posted on a single page so there must be 215.7 pages total which is not possible, therefore we will be limiting ourselves to even 200 pages.

```
In [3]: link_list = []
        original_link = ['https://www.seek.com.au/software-developer-jobs']
        middle = ['?page=']
        for i in range(200):
            new =(original_link[0]+middle[0]+str(i))
            link_list.append(new)
```

The following code gives us a list containing all the links to different pages which we will feed to the extractor program directly. A part of the list is shown below.

```
In [4]: link_list
        'https://www.seek.com.au/software-developer-jobs?page=181',
        'https://www.seek.com.au/software-developer-jobs?page=182',
        'https://www.seek.com.au/software-developer-jobs?page=183',
        'https://www.seek.com.au/software-developer-jobs?page=184',
        'https://www.seek.com.au/software-developer-jobs?page=185',
        'https://www.seek.com.au/software-developer-jobs?page=186',
        'https://www.seek.com.au/software-developer-jobs?page=187',
        'https://www.seek.com.au/software-developer-jobs?page=188',
        'https://www.seek.com.au/software-developer-jobs?page=189',
        'https://www.seek.com.au/software-developer-jobs?page=190',
        'https://www.seek.com.au/software-developer-jobs?page=191',
```

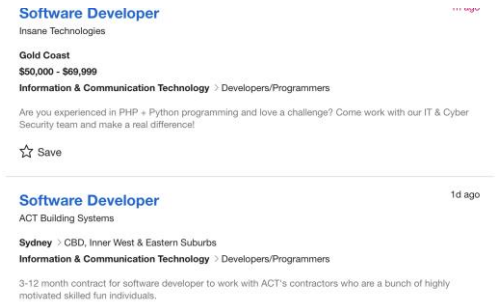
Now we will start with our main program which will go through each one these links, open every job profile posted and extract the job description.
We start with calling our main libraries which we will be using as shown.

```
In [5]: from bs4 import BeautifulSoup as soup
        from urllib.request import urlopen as uReq
```

Beautiful Soup is a Python library for pulling data out of HTML and XML files[8] while the urllib.request module defines functions and classes which help in opening URLs (mostly HTTP) in a complex world — basic and digest authentication, redirections, cookies and more[11].

```
In [9]: for url in link_list:
            uClient_all_jobs = uReq(url)
            page_html_all_jobs = uClient_all_jobs.read()
            page_soup_all = soup(page_html_all_jobs, "html.parser")
```

The uReq will be one calling the links we stored earlier in link_list. It will make a connection with the respective link and the .read() function will store all the html data stored in that particular page. Then our main function which is soup() from Beautiful Soup which will essentially clean the html code we have stored and save it in a form on which our code can perform different html extractions.



This is what a structure of the page looks like containing different job postings. Each of the Headings of the job contains a link which leads us to the job description. So we have to look for such links, open it and extract the job description.
Inspection the element of each of these link I found that each of the headings is under an <a> tag having the same class and data-automation type. Because this structure is straightforward to write the code.

```
for data in page_soup_all.findAll('a',{'data-automation':'jobTitle'}):
    link = data.get('href')
```

In this we are looking the <a> tags with data-automation type – jobTitle and directly extracting the link (“href” is basically used in html to store the link to another page) stored in it.
Now the link we have extracted from this is not complete i.e. “/job/50075763?type=standard#searchRequestToken=d6a32227-b39c-4fd1-87a0-a62b2b49c419".
It has to be concatenated to the original link for the website to make it accessible. This is shown in the figure below.

```
website = ['https://www.seek.com.au']
new_link = website[0]+link
```

This will give us all the links for the jobs posted in single page and we are looping through all the pages, basically we have all the separate links for the job postings for software developers. A part of it is shown below.

```
https://www.seek.com.au/job/50036571?type=standard#searchRequestToken=9bb14cc8-2e8d-4336-8aba-103fe55cd411
https://www.seek.com.au/job/50013812?type=standard#searchRequestToken=9bb14cc8-2e8d-4336-8aba-103fe55cd411
https://www.seek.com.au/job/50049454?type=standout#searchRequestToken=9bb14cc8-2e8d-4336-8aba-103fe55cd411
https://www.seek.com.au/job/50060378?type=standard#searchRequestToken=9bb14cc8-2e8d-4336-8aba-103fe55cd411
https://www.seek.com.au/job/41355622?type=standard#searchRequestToken=9bb14cc8-2e8d-4336-8aba-103fe55cd411
https://www.seek.com.au/job/50012974?type=standout#searchRequestToken=9bb14cc8-2e8d-4336-8aba-103fe55cd411
https://www.seek.com.au/job/41360533?type=standard#searchRequestToken=9bb14cc8-2e8d-4336-8aba-103fe55cd411
https://www.seek.com.au/job/41360535?type=standard#searchRequestToken=9bb14cc8-2e8d-4336-8aba-103fe55cd411
https://www.seek.com.au/job/50075192?type=standout#searchRequestToken=9bb14cc8-2e8d-4336-8aba-103fe55cd411
https://www.seek.com.au/job/50024062?type=standout#searchRequestToken=9bb14cc8-2e8d-4336-8aba-103fe55cd411
```

Now that we have all the separate links for job postings, we just have to open each link and look for job description and then extract it.

```
uClient = uReq(new_link)
specific_course_html = uClient.read()
page_soup_specific = soup(specific_course_html, "html.parser")
```

The above code again makes a request to open the link as we did earlier with urllib.request, reads all the html information from that link and then we parsing the html data with help of BeautifulSoup. Now we have the entire html code for all the job postings in a page.
Each of the job listing looks something like,

Candidates require the following knowledge, skills and experience:

- Excellent understanding of **PHP, JavaScript, Postgres and MySQL**
- Excellent understanding of **Python and Perl** programming languages
- Excellent understanding of working with **RESTful API** endpoints
- Good understanding of **Regular Expressions** (RegEx)
- Good understanding of **networking fundamentals** (TCP/IP, OSI layer, etc)
- Good understanding of **operating systems** (Windows, Linux, MacOS)
- High degree of **empathy, diplomacy and confidentiality** at all times
- **Excellent organisational and communication skills (verbal and written) required**
- Proactive ability to manage own time effectively

And analyzing the html code for the description, I found that the entire job description lies under a tag with "*data-automation*" of the type "*mobileTemplate*". So we are going to look for these tag in our html code and extract the entire text (which is the job description in our case) as follows.

```
for job_over in page_soup_specific.findAll('div',{'data-automation':'mobileTemplate'}):
    overview = job_over.text
    over.append(overview)
```

This code is taking the entire job description as text in overview. Then we are appending all the text in a list called over[]. This will store the different job description in list called over which we can use later for analysis.
We will save the same thing in a data frame for better visualization for what we have collected.

```
dfObj = pd.DataFrame(columns=['requirements'])
for job_over in page_soup_specific.findAll('div',{'data-automation':'mobileTemplate'}):
    overview = job_over.text
    over.append(overview)
    dfObj = dfObj.append({'requirements': overview }, ignore_index=True)
```

Here we are initializing our data frame and then appending all the job description in it. Then exporting the all the text in a csv file we get.

```
In [37]: dfObj.to_csv('overview.csv')
```

	requirements
2	MooGoo is an Australian-made skin care company based in the sunny Gold Coast. Founded in 2005, MooGoo
3	We are looking for talented Software Engineers in Support to provide our customers an outstanding customer s
4	Who is Zip? People at the ASX call us a 'Unicorn Fintech Start-up'.More than 1.8 million Australian shoppers ca
5	Are you a GIS Software Engineer with experience in designing and developing of a relational geodatabase (Pos
6	Monday-FridayFull-time About Easygo Easygo Solutions Pty Limited is an Australian company that is looking to
7	One of our premium clients based in Melbourne is looking for a highly experienced and committed individual fo

Clicking on any of the rows we get the entire description in text form as shown.

6	Monday-FridayFull-time About Easygo Easygo Solutions Pty Limited is an Australian company that is lo
7	One of our premium clients based in Melbourne is looking for a highly experienced and committed indiv
8	and Skills Required:· Minimum 2yrs+ hands on development experience on Node JS, Core Java, Spring Mandatory Skills:· Node JS, Java, Microservices, Springboot Developer, Splunk experience Good to Ha
9	Pen testing, Vul scans, AWS services such as Lambda, DynamoDB, CloudFormation, ECS, Secure code have delivered world class IT and business consulting services across a wide range of industries. As a r
10	committed, can-do individuals have always been the cornerstone of this dynamic business. It is expect

4.2.2 For RQ3

For this we will directly use survey conducted by James J. Cappel (2002)[9], which will help us understand what are large organization looking for in terms of non-technical skills. His survey was completed on an anonymous basis, with a response rate of 36% from organizations. Of all the organizations responded 40.7% represented Computer/Computer Services, Of these 77% were valued at more than $500 million with 66% having a staff of more than 150 each [9]. The survey had employers rate expected skill requirement and provide actual data of the skills the employee had. The following table 1 is taken from James J. Cappel paper directly [9], which gives an overview of the companies involved and percentage response from them.

TABLE 1
Sample Characteristics

	Percentage
Industry	
Computer/computer services/IT consulting	40.7%
Healthcare, pharmaceuticals	14.8%
Financial services, insurance	11.1%
Aerospace/defense	3.7%
Manufacturing	3.7%
Publishing, printing	3.7%
Chemical, oil and gas	3.7%
Utilities	3.7%
Other	14.8%
Company size (in annual revenue)	
Less than $100 million	11.1%
$100-499 million	11.1%
$500-999 million	14.8%
$1-3.9 billion	7.4%
$4-9.9 billion	11.1%
$10 billion or more	33.3%
No response	11.1%

Although this survey also answers other questions like which programming language was in more demand than others, we won't looking at that data as this survey represents data from 2002, and programming landscape is now somewhat different, so we will be only looking into the Non- Technical skills data which belongs in the Table 4 of the paper, as shown below.

TABLE 4
Non-Technical Skills

	Expected	Actual	Difference	P
Skills and abilities:				
Ability to learn	4.78	4.11	0.67	.0000***
Teamwork	4.37	3.81	0.56	.0001***
Problem solving	4.07	3.15	0.92	.0000***
Written communications	4.04	3.19	0.85	.0000***
Oral communications	4.04	3.07	0.97	.0000***
Time management	3.81	3.26	0.55	.0003***
Ability to work under pressure	3.77	3.08	0.69	.0000***
Ability to apply IT to business problems	3.70	2.96	0.74	.0001***
Ability to resolve conflict professionally	3.63	2.93	0.70	.0005***
Change management	3.00	2.26	0.74	.0001***
Personal qualities:				
Motivation to work	4.63	3.74	0.89	.0000***
Initiative	4.59	3.70	0.89	.0000***
Professional ethics	4.37	3.85	0.52	.0039**
Attention to detail	4.30	3.26	1.04	.0000***
Persistence	4.07	3.26	0.81	.0000***
Patience	3.81	3.19	0.62	.0003***
Leadership	3.59	3.00	0.59	.0001***
Maturity	3.56	3.15	0.41	.0133*

*p<=.05, **p<=.01, ***p<=.001

4.3 Data Analysis

4.3.1 RQ1

For RQ1, we will be looking for keywords from our list of job description. The list of keywords will be based upon some of the most common languages used in today's IT industry and randomly viewing job listings and checking the requirements. It will include languages such as JavaScript, Python , HTML, SQL, Java, C#, Typescript, C, C++, Ruby, Go, Swift etc.
We will start with overview we have collected from the job postings, as we know it is stored as text in our "over" variable. We will split each of the job description by spaces and count the instances the words are repeated from our predefined list.

```
for all_info in over:
    total_words = all_info.split()
```

We have defined the keywords as follows,

```
languages = ['JavaScript','HTML,','CSS','SQL','Python','Java','Bash',
             'Shell','C#','PHP','C++','TypeScript','C',
            'Ruby','Go','Assembly','Swift','AWS','Docker','MySQL','React','Node.js']
```

Now we will look through all the words we have split and compare with our keywords and it will only store the words in overview which are present in our keywords. We will define a function which will do this for us. [16]

```
def language_keywords(total_words, languages):
    return [w for w in total_words if w in languages]
```

Now we will call this function as follows,

```
wordlist = language_keywords(total_words,languages)
dictionary = wordListToFreqDict(wordlist)
sorteddict = sortFreqDict(dictionary)
```

The other two functions in this are just counting the frequency and then sorting them out in descending manner.

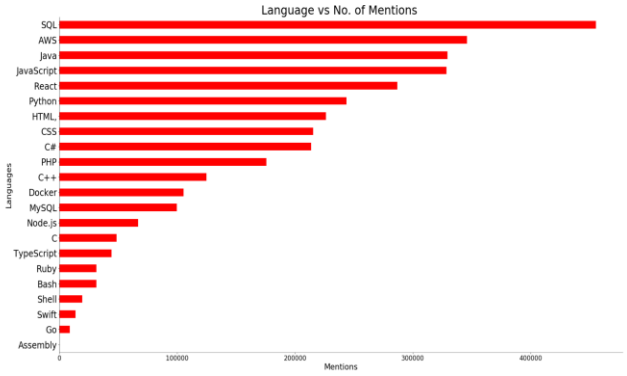
Afterwards we will be using counter from collections so that it will add the same key-value pairs in our dictionary and give a final result.

```
for key_value_sorted in sorteddict:
    counter.update(key_value_sorted)
result = dict(counter)
```

The above code if we run for around 210 jobs postings (Just for a reference so that computing time is low) , we get the following result,

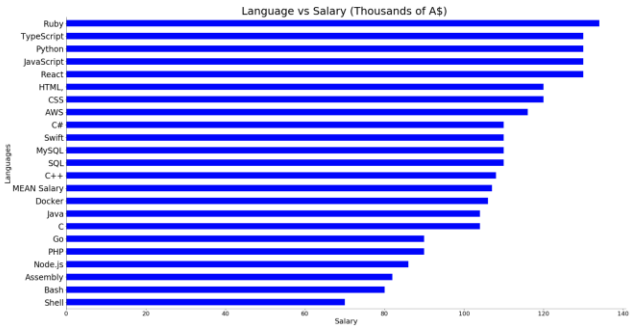
```
'CSS': 22643, 'SQL': 19792, 'JavaScript': 19692, 'HTML,': 18925, 'AWS': 15917, 'PHP':
14618, 'Java': 14342, 'C#': 12411, 'React': 11776, 'MySQL': 9892,
'Python': 7518, 'Docker': 5011
'C++': 4574, 'Bash': 933, 'TypeScript': 826, 'C': 810, 'Ruby': 571, 'Node.js': 511,
'Go': 318, 'Shell': 1 63 'Swift': 15})
```

The following result shows the total number of time our keywords were mentioned in the job posting in descending manner.
Some of the keywords are in ten's of thousands as sometimes the requirements are mentioned more than once in a job listing and some job listings require more than one languages as the job requirements. This should still reflect the demand of a particular language over another.
For 210 job listing it took around 1 minute of computing time, so for running our code through 4000 jobs, it would require around 20 minutes of computing time. The final result for better visualization will be shown as graph.



4.3.2 RQ2

For this research question, we are going to analyze different job listings at random to check the expected wages given and map it the respective programming skill requirement. The "*new_link*" we generated in section 4.2.1 will help us achieve this. We randomly analyzed about 100 different job postings saved their result. The final result is shown below, with mean salary also shown and plotted with rest of them.



Average salary for Software developer being "*107.695K*"

4.3.3 RQ3

The data we are using to analyze is a survey conducted by James J. Cappel[9] and it had employers rate expected skill requirement and provide actual data of the skills the employee had.

TABLE 4
Non-Technical Skills

	Expected	Actual	Difference	P
Skills and abilities:				
Ability to learn	4.78	4.11	0.67	.0000***
Teamwork	4.37	3.81	0.56	.0001***
Problem solving	4.07	3.15	0.92	.0000***
Written communications	4.04	3.19	0.85	.0000***
Oral communications	4.04	3.07	0.97	.0000***
Time management	3.81	3.26	0.55	.0003***
Ability to work under pressure	3.77	3.08	0.69	.0000***
Ability to apply IT to business problems	3.70	2.96	0.74	.0001***
Ability to resolve conflict professionally	3.63	2.93	0.70	.0005***
Change management	3.00	2.26	0.74	.0001***
Personal qualities:				
Motivation to work	4.63	3.74	0.89	.0000***
Initiative	4.59	3.70	0.89	.0000***
Professional ethics	4.37	3.85	0.52	.0039**
Attention to detail	4.30	3.26	1.04	.0000***
Persistence	4.07	3.26	0.81	.0000***
Patience	3.81	3.19	0.62	.0003***
Leadership	3.59	3.00	0.59	.0001***
Maturity	3.56	3.15	0.41	.0133*

*p<=.05, **p<=.01, ***p<=.001

In the following table skills are rated from highest to lower with each skills with rating higher than four on a five-point scale. Some of the skills are "*ability to learn (4.78), Teamwork (4.37), Oral communication (4.04)*", some other moderately important ratings were also considered such as "*Time Management (3.81), Apply IT to Business Problems (3.7)*"[9].

We can clearly see from the survey of the employers that greatest mean difference for skills and abilities of the "expected" and "actual" occurs for "*oral communication (0.97), problem solving (0.92) , and written communication (0.85)*"[9].One more important thing to note in this is that "*difference between expected and actual of every non-technical skill was greater than 0.50 and these differences were all statistically significant*[9]".

For the personal qualities section, some of the top rated qualities are "*motivation to work (4.63), attention to detail (4.30) etc*" and largest mean difference between the "expected" and "actual" values occurred for "*attention to detail (1.04), motivation to work (0.89) and initiative (.89)"[9]*. In personal qualities section as well "*difference for every quality was greater than 0.50 and these differences were all statistically significant*"[9].

5 FINDINGS & DISCUSSIONS

5.1 "What will be the top skill requirement in the near future, in terms of the programming skills".

We can clearly the see from the graph plotted which of the programming skills are required right now in Australia, with SQL taking the top spot. One thing to note in the graph is that SQL and MySQL are separate as SQL is a query language and MySQL is a database software. It used "SQL" language to query the database. So one can predict if we combine these two the requirement for SQL language will be even higher.

Same can be said for Node.js and React as both are subsets of JavaScript i.e. you need to understand JavaScript to work on Node.js and React.

We have divided these into separated sections as requirements of some environment were higher than other and it gives us a better understanding of which skill is more in demand and indicates more jobs.

For a reference we checked other articles working of this specific problem, We picked a popular article on medium[12] and we found out that the top spot is taken by Python, followed by JavaScript, Java, C#, and C, which indicates quite interesting results as Python is quite lower on our list.

This shows that in-demand programming skills are different for different regions as this article doesn't specify the basis of their data or region. What might be in demand skill for a specific country might be totally different in other country as shown in our analysis.

5.2 "**How are these skills related to the wages**?".

According to the data from "living Australia"[13], The average salary for an Australian (2018 Quarter-2) is **A$ 86,436.**

We can clearly see that a job in Software development will earn you an average salary of around **A$107,695** which is an **24.59%** increase. This shows that a job in Software Development pays significantly higher than the average. One surprising outcome of this was highest paid programming skill turned out to be Ruby which would fetch you an average of A$135k per annum. Maybe this might be because there aren't many programmers coding in Ruby as a result it pays so much. Followed by Ruby were results matched by RQ1 finding i.e. JavaScript and Python being close second averaging around A$130k per annum.

5.3 "Importance of non-technical skills compared to technical skills".

Most people specially in Software Engineering never focus on non-technical skills, they are mostly focused on the technical aspect which is important but the non-technical aspect should not be ignored. From the survey of the employers it is clearly visible that they don't get what they expect in terms of personal qualities. They is a disparity here between employers and employees, for example in "oral communication" which saw the highest difference, which should be concerning as any knowledge is not useful in a workspace if you are not able to communicate it with people around you. Also a quality which scored higher difference between expected and actual was "attention to detail" and

"motivation to work" which scored 1.04 and .89 respectively. This suggest that most people are working the jobs they are interested in, or their work has come to a saturation level. Therefore it very important to find something you can do everyday without a reduction in effort.
People talk about non-technical skills being important in a workspace but they don't generally provide data to support their claim. But with this survey provides hard evidence from the employers directly that non-technical skills are as important as technical skills for an individual to have a successful career.

6 THREATS TO VALIDITY

The research done in this paper isn't perfect, from the research proposal to actually doing the research a lot of changes had to be made. Some of the potential threats to validity are,

Firstly I was supposed to scrape data from three services Seek, LinkedIn and Indeed for a total job listings of around 12,000 and I was only able to scrape data of about 4400 jobs listings. Scraping data from LinkedIn would have validated the results of this study more as it a more trusted website on a global platform but LinkedIn doesn't allow third party to scrape data from their website or they have a firewall blocking the way I collected data for this research, which meant I had to drop data collection from LinkedIn. Also we surveyed some of the job listings from Indeed and found out that much of the job listing on this platform and on Seek were same which would have led to a lot of repetition, so I decided to drop Indeed as well. This led to our dataset being considerably reduced from what was originally planned. Over 4000 jobs are still good data set but more data would have resulted in more insightful results.

The main result for RQ1, i.e. the graph we plotted between Language and Number of Mentions isn't exact. The number of mention sections have values in hundreds of thousands which sounds impossible when you consider that there were only around 4400 jobs. The error in this lies in my approach of how I calculated number of mentions, Instead of calculating language mentioned in a job listing only once, we are calculating all the mentions of a particular job in a listing. This lead to inflated numbers for mentions as most of the job listings were continuously repeating the keywords we set, and the number kept going up. But it would still give us an idea about in demand languages as more mentions overall would suggest more demand, even though the exact number is inflated.
Also a potential threat to validity for RQ1 is based on an article by Wall Street Journal[15], which talks about how the technology sector has lost record number of jobs due to Covid-19. Although the article is based out of United States, the reduction in jobs is something which is resonated around the world and could potentially affect results for this questions.

In the result for RQ2, I was supposed to collect expected salary given in a particular job listing and map it onto the required languages for that listing. But this data on job listing was displayed with the use of JavaScript rather than being an HTML code. This led to me not being able scrape data using beautiful soup as beautiful soup doesn't parse JavaScript. Because of this I had to manually look into random links of job listings and add the expected salary to languages and plot it. That's why graph for languages vs average salary isn't as accurate as the graph for RQ1 which had significantly bigger data set as it was automated and RQ2 was manual.

The paper by James J. Cappel[9], which I have used to answer RQ3 was published in the year 2002 which makes it 18 years ago, which in the IT industry is just plain outdated. As we know that the programming landscape have changed in last 18 years, so can be the qualities which the employers seek. Therefore the results for RQ3 has to be taken with a grain of salt as the survey on which it is based on is really old. Also we are using a result from a survey and in surveys there's always a risk of losing honesty. But this survey was completed on an anonymous basis which mitigates this risk.
I wanted to do my own survey from employers based in Australia to answer this question but to conduct a survey for a research you need to get approved from the Ethics board in the University which takes around 2 months, which is also a timeframe to complete this research paper therefore making it impossible. If it was possible the answer to RQ3 would have been much more accurate and would have reflected todays requirement of non-technical skills by the employers.

7 RELATED WORK

As this is a localized study i.e. I have only analyzed data based in Australia, there is not much related work but there are studies on this topic done in other countries.
Like the paper by James J. Cappel[9], in which he surveyed large list of employers from the IT industry in the United States. This survey addressed employers perceptions of importance of programming languages and various other technical and non-technical skills. And also how students can better prepare themselves for the profession prior to graduation. The main difference here is the dataset used by Cappel[9], which belongs to small numbers of employers compared to my paper, which had requirements from thousands of companies. Obviously at the time the technology was simply not available to do this research like I have done. Still the survey of the employers for the non-technical requirements was used as you still cannot find such important data with the data backing it.
There are some other research papers which answers some of these question asked like developing small number of specific skills or whether you should learn broad range of skill [2], it also talks about which skill leads to highest wage.
Carnevale et al.[3] talks about "shift from the industrial era to the post-industrial era of the knowledge economy" and how this resulted new skill requirements from companies, technical skills augmented with growing need for non-technical skills like

"learning, reasoning, communicating, general problem-solving skills and behavioral skills" which is the basis for RQ3
Decker, MR et al.[14] talks about the disparity between skill set desired by the employers and skills employees sought how this has consequences over the quality of students graduating from universities.

8 CONCLUSIONS

In conclusion there's appears to be productive result from the data I have collected which could help people specially students or people who are thinking of taking software development as their career path. As the result is entirely based on data collected from Australian jobs postings, this could help the local community understand the skill requirement with respect to their country, In our case SQL being in demand the most. In addition it will give them additional information regarding the wages related to that skill, average salary for mentioned skills is around A$107K per year and illustrate the importance of non-technical skills in the workspace backed by the data provided by the employers with "oral communication" and "attention to detail" are some examples which employers seek the most in employees.
The results hopefully helps people make better choices in their career going forward.

9 FUTURE WORK

I think for each the three proposed question, the research can be continued and provide us with more accurate results. First being the data collection for RQ1, rather than just collecting data from single platform, one can collect data from all the platforms available to that region, avoid repetition and then analyze depending on our requirement. This will give probably the most accurate result for our question can be useful to thousands of students and people, even government can use this data to maybe improve funding in public schools to improve the learning of the skills which are most in demand.
Secondly, for RQ2 rather than just randomly analyzing job postings to collect data for wages, a process must be developed that does this for all the listings, which would give us a more accurate result.
Thirdly a updated survey needs to be conducted for RQ3 rather just using an old survey. Conducting the survey at this time from the employers present locally would make results more related to current time.
All the research questions would give us the most comprehensive study about skills requirements in the software development.

10 REFERENCES

[1] Article on programming languages from Wikipedia as accessed on 14th June 15, 2020 https://en.wikipedia.org/wiki/List_of_programming_languages.
[2] Rotundo, M & Sackett, PR 2004, 'Specific versus general skills and abilities: A job level examination of relationships with wage', Journal of Occupational and Organizational Psychology, vol. 77, no. 2, pp. 127–148.
[3] Carnevale, AP & Smith, N 2013, 'Workplace basics: the skills employees need and employers want', Human Resource Development International, vol. 16, no. 5, pp. 491–501.
[4] Article on seek from Wikipedia as accessed on 14th June 15, 2020 https://en.wikipedia.org/wiki/Seek_Limited
[5] Seek as accessed on 14th June 15, 2020 https://www.seek.com.au/about/
[6] Article on Anaconda from Wikipedia as accessed on 14th June 15, 2020 https://en.wikipedia.org/wiki/Anaconda_(Python_distribution)
[7] Article on Jupyter Notebook from Wikipedia as accessed on 14th June 15, 2020 https://en.wikipedia.org/wiki/Project_Jupyter#Jupyter_Notebook
[8] BeautifulSoup Documentation on Crummy as accessed on 14th June 15, 2020 https://www.crummy.com/software/BeautifulSoup/bs4/doc/
[9] Cappel, JJ 2002, 'Entry-Level is Job Skills: A Survey of Employers', Journal of Computer Information Systems, vol. 42, no. 2, pp. 76–82.
[10] Seek, accessed on June 11, 2020, https://www.seek.com.au/software-developer-jobs
[11] Pythons forum for URL requests as accessed on 14th June 15, 2020 https://docs.python.org/3/library/urllib.request.html
[12] Article on Programming Languages as accessed on 14th June 15, 2020 https://towardsdatascience.com/top-10-in-demand-programming-languages-to-learn-in-2020-4462eb7d8d3e
[13] Living Australia as accessed on 14th June 15, 2020 https://www.livingin-australia.com/salaries-australia/
[14] Decker, MR, Bronson, NW, Greenberg, CC, Dolan, JP, Kent, KC & Hunter, JG 2013, 'The General Surgery Job Market: Analysis of Current Demand for General Surgeons and Their Specialized Skills', Journal of the American College of Surgeons, vol. 217, no. 6, pp. 1133–1139.
[15] Wall street Journal as accessed on 16th June 15, 2020 https://www.wsj.com/articles/technology-sector-shed-record-number-of-jobs-in-april-11588974127
[16] Code for Data Analysis used, accessed on 14th June 15, 2020 https://programminghistorian.org/en/lessons/counting-frequencies