HelpJobs – understand jobs better

<https://jordancfq.shinyapps.io/tfgh/>

IS428 AY2020/21 Term 2 G1 Group 2

Chen Fang Qi

Lim Yin ShanKee Cheng Heng

Thinesh S/o KalaichelvanLee Min Hui

*Abstract*— Providing a comprehensive overview of Singapore’s job market through a self-help tool

# Introduction

Unemployment is a prevalent issue around the world, and Singapore too, is not spared from it.

The unemployment rate in [Singapore](https://www.statista.com/topics/2721/singapore/) rose to around 4.38 percent in 2020 from 4.11 percent in the year before. With the rise of Covid-19, companies are shutting down and employees are getting retrenched.

In addition, Covid-19 is not the only reason for the increasing unemployment rates (cyclical), structural unemployment also contributed to it, which means there are changes in skill set requirements based on the latest trends within the market. With companies shutting their services, we see huge changes in the job requirements as each industry pivots rapidly to adopt new innovative practices. You see companies requiring multi-faceted workers rather than workers from a single discipline. E.g. Developers who are artistic - Virtual Reality scenarios and even videographers who could code - Mixed Reality productions.  
  
We face an extremely real problem today. It is increasingly difficult to understand the job market as industry standards constantly shift in our fast-paced society. What was new last year is old news this year. Popular industries quickly become graveyards, while new up and coming industries constantly cause disruption and fight to remain relevant in the market. All of this makes it harder to understand the job market and what the common skill sets that are being sought after. It is extremely difficult to obtain advice from those who went through the industry as they would have probably been outdated and information is no longer accurate. The best way to aggregate and make actionable insights is by actively surveying the latest information and reviewing recent datasets.

Lastly, because of the lack of jobs and shifting industry standards, it has become harder to assess our own value. We see more people asking themselves questions such as:

- “Am I getting underpaid?”

- “Should I be asking for more salary?”

- “Am I too demanding in my requests?”

These are just a few of the numerous questions that will commonly appear as we navigate around the workforce.

Despite the current efforts by job listings employment agencies creating filters and centralised platforms for interested parties to narrow down to their ideal jobs, they are still lacking in many areas such as more in-depth job comparisons, higher salary transparency and more - which hope to solve through visual analytics today.

By creating an r-application that is user friendly and attractive to end-users, we hope to allow complex datasets to be self-served by user groups to reach their own conclusions and analysis on the publicly available information. As a non-biased source of consolidated information, this will help stakeholders to arrive at better decision makings through validated actionable insights.

# METHODOLOGY

## **Data Collection & Pre-Processing**

The programming language used to scrape the data was Python. Specifically, the modules used were: Requests and Pandas.

To scrape the data, there are a few known ways to do so:

Using Selenium/Request and processing the text scraped with BeautifulSoup

However, if we do it in this method, there would be a lot of unnecessary resultant noise in the form of CSS and HTML elements. This will result in the need to do quite a heavy bit of post processing to make the data usable for analysis.

An alternative way was to get the data directly via an API endpoint made available due to the various services in use at the website. Although this API isn't publicly available, this endpoint can be found by applying a Man in the Middle approach to see how the data is being populated and therefore making calculated assumptions to eventually derive the data endpoint.

We ultimately opted to go with the Man in the Middle approach as the data was much cleaner and can be immediately utilised for processing rather than having to run multiple corrections to clean it to a usable state.

The endpoint we used to scrape this was: <https://api.mycareersfuture.gov.sg/v2/jobs>

With this endpoint, it returns the data in a JSON format. After accessing the endpoint, we were able to retrieve and process the data. We filtered out the Job Types such as Internship and Part Time as we realised that it was creating a bunch of unnecessary noise due to unavailability of data. We observe quite a number of $0 salary as such jobs were more flexible in the pay attractivity.

Throughout the job listings, there was a Minimum Salary and Maximum Salary dataset that was available. However, we realised that the spread could be extremely big and hence our introduction of Adjusted Salary column as an attempt to normalize the information.

We performed the following formula to make the adjustment of Salary based on the various factors involved.

if ((Min Salary + Max Salary) \* 0.4) < Adjusted Salary:

          adjusted salary = adjusted salary

else:

        adjusted salary = min salary

Moreover, to avoid potential anomalies, we have filtered jobs that pay more than $10,000 as our end audience is intended to be targeted at fresh graduates who are unlikely to fetch such a high starting salary.

Within our intended charts, we had to plot an interactive map. This required coordinates of each company to be available. The only data we could find was in Lat and Lng whereas tmap uses X and Y. So we had to find an alternative to tmap to tackle this limitation.

We found Leaflet, which was able to visualize based on Lat and Lng data.

Since Lat and Lng data was also missing in the original endpoint, we had to dig around to use another man in the middle approach to locate an alternative endpoint which was holding Lat and Lng data points. We managed to leverage on this endpoint to fix our issue:

https://api.mycareersfuture.gov.sg/v2/jobs/<job uuid>

This endpoint not only has the Lat and Lng data but it also has the Skills that the employers are explicitly looking for as mentioned within the job listing.

To process the Skills list easily, we executed String Manipulation via Python where we split the Skill list and only retained the unique ones and the ones that have a standard capitalization as we noticed that this approach is more consistent and produces a result of better quality for our eventual data processing & analytics.

This approach was also used to scrape data from another Platform, Glints. We faced similar data gathering challenges. However, a unique pain point was that Glints has a different bin for industry so we couldn't show the salary distribution across all the industry and only those that both Glints and MyCareersFuture have in common.

We managed to include the Glints dataset within our application to quickly demonstrate and illustrate how we can compare various platforms and even place salaries posted by the employers from these platforms side by side for analysis. We were also able to detect abnormalities and make recommendations as well as cross check assumptions to validate their accuracy. By having a base layer, we were able to create a benchmark that would now whole greater weight due to its larger data points.

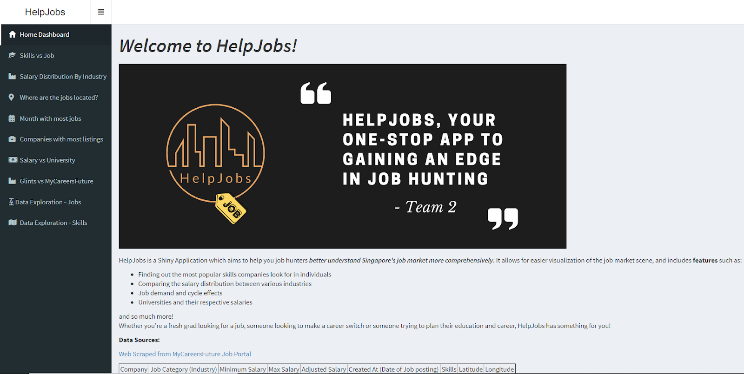
MyCareersFutures is a portal that aims to provide Singapore Citizens and Permanent Residents with a fast and smart job search service to match them with relevant jobs, based on their skills and competencies. The portal was developed by Workforce Singapore, in partnership with the Government Technology Agency. The portal enables jobseekers to be more aware of the skills they possess, and connects them to relevant jobs based on their current skills and competencies. It also highlights jobs which are eligible for Government support through WSG’s Adapt and Grow Initiative.

We made the decision where MyCareersFuture would be the most stable dataset as it is regulated and screened by the government. There should be a Standard Operating Procedure put in place which would have been circulated so that the data being collected would properly align with the requirements and reporting.  We can also guarantee that there would be an adequate number of data being constantly added as time passed by due to government initiatives and even government directions for companies that were related to the government sector.

# VISUALISATION

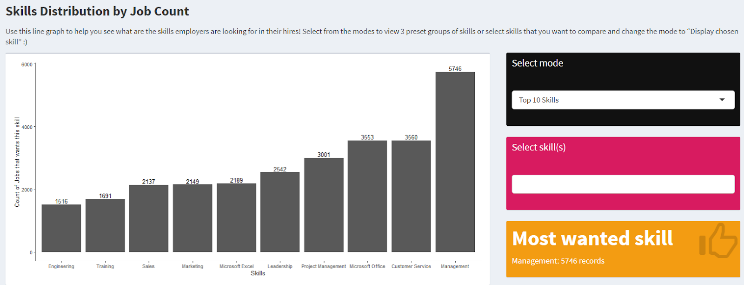
The visualisations were all performed in R and segmented into 10 tabs.

## **Dashboard Home Page**



The dashboard home page gives a brief introduction to what HelpJobs can provide for users, including its features and information about the datasets used. There is also a left sidebar which users can click on its various tabs to dive into the features of HelpJobs.

## **Skills vs Jobs**

****

The “Skills vs Jobs” visualisation aims to help the user identify which skills are the most popular based on the number of jobs requesting those skills.

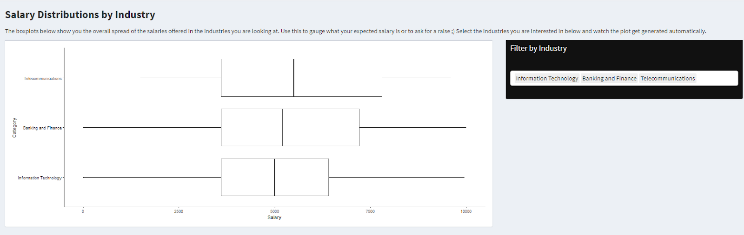
We used bar charts to clearly visualise the differences in popularity between the skills. Having the bars side by side, with the job count visible as well, makes it easier for the user to make comparisons.

Our visualisation has 4 modes as well. The first and default mode is “Top 10 Skills”. This mode displays the top 10 most popular skills overall. The second mode “Display chosen skill”. This mode allows the user to input certain skills they want to see. They can type in or select from a drop down list. They will then see a custom chart on the popularity of the skills they are interested in. The third and fourth modes are “Programming Skills” and “Business Development Skills”. As we were working on the project, we realised there are certain skills that could be grouped together. That is how we created the programming and business development groups. Selecting either option will display a visualisation of a few, key skills that are important for those selected sectors.

The chart allows users to clearly see and compare the demand for different skills. THis allows them to better value themselves when negotiating pay or salary raise, if they know their skills are in demand. It also helps them decide which new skills to pick up, based on its popularity in the job market.

We discovered through the visualisation that Management was one of the most sought after skills, that appeared on 25% of job listings. This could be a sign of the shift in employment strategies, where employers no longer just want worker-bees who follow instructions. Instead they want each individual worker to be able to manage their own projects and make decisions.

## **Salary Distribution by Industry**

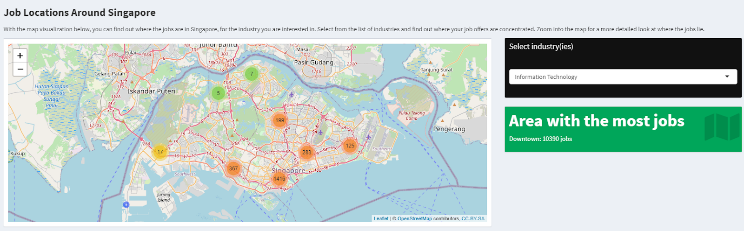
****

The “Salary Distribution by Industry” aims to give users an overview of the salary distribution across the various industries in Singapore. Users are able to use this chart to understand where their current salary stands in the industry, if they are underpaid and even how much they should ask for especially if they are taking a new role. This visualisation also aims to find out which industry has the highest compensation.

To do so, we used a box-plot which shows the range, the interquartile range, the bottom 25% the median, the top 25% as well as the outliers . Users are also able to compare the salary between industries in the same chart by selecting or typing the industry in the drop down list on the right. This comparison between industries will be extremely useful for working adults who would like to work in a different industry, for students to understand the industry they would like to work in as well as the Singapore government on planning the future.

Through our visualisation, we found that the telecommunication industry has the highest median salary across all other industries while hospitality is the lowest. One reason why the hospitality industry is the lowest might be due to the low barrier of entry into the industry. Next, we also found that around 75% of the industries have their median within the $2,500 to $3,500 rang. Hence, we can conclude that the salary of a typical person working in Singapore will be between $2,500 and $3,500.

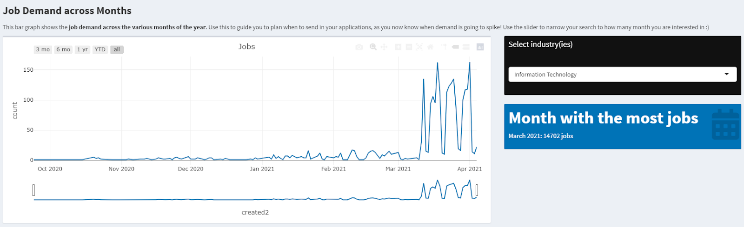
## **Where are jobs located**

****

This map chart shows the distribution of the job’s location in Singapore. Through this visualisation, we aim to find if there are any patterns in the location of the jobs in the different industries. Potential job seekers can use this chart to understand what kind of jobs are available around where he/she is currently living in, or even rent a place that is closer to the industry he is going to work in.

To create this visualisation, we used the package ‘Leaflet’ and we added the markers on the map using the location of each job. We also added a drop-down list on the right through shiny to filter the different industries. Through the map chart, we found that the downtown area has the highest concentration of jobs listed on the platform. We also found that certain industries have very distinct patterns in where they tend to be located. For instance, the insurance industry is mostly located in the central business district area, the education and training industry are located all over Singapore which is according to what the government planned.

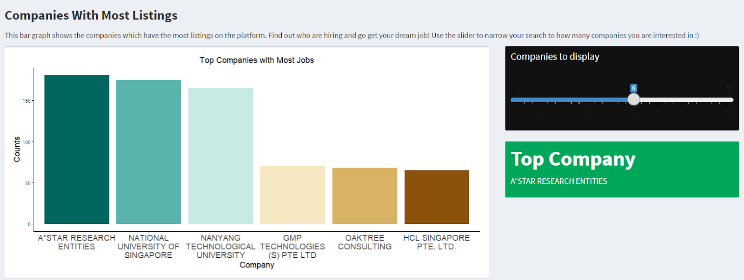
## **Month with most Jobs**

****

In this visualisation, we want to understand the seasonality of the job search especially if there are any correlation among the different industries. Also, this chart can help users to understand if a particular industry is saturated and full. Through this visualisation, we want users to understand when is the best time for them to begin their job search and whether different industries have different hiring seasons.

We decided on using a time series line chart as it emphasises any patterns and trends present in the dataset. Users are also able to use the slider at the bottom to zoom in to the time period they are interested in. A drop-down list was also added to filter according to the various industries. Through this visualisation, we realised that there is a huge flaw in our dataset as there is a huge spike in the number of job postings in March. One reason for this might be that there was a bulk upload exercise on the MyCareerFuture platform but a more probable reason is that the majority of the jobs posted before March have been filled and hence delisted on the platform. From this, we can infer that jobs postings tend to be filled within a month. We also found that the Information Technology industry has the highest number of postings and can conclude that the industry is still expanding.

## **Companies with most listings**

****

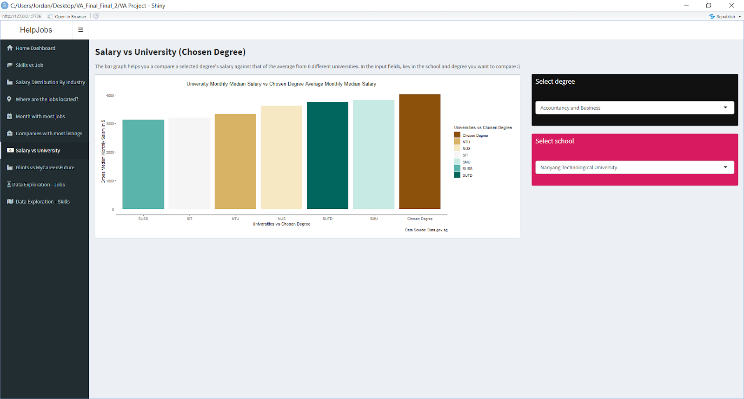
The “Companies with most listings” aims to visualise which companies have the most job offers and allow users to compare the number of job offerings amongst the top few.

We used a bar chart as its clean and simple design makes it easy for the user to quickly grasp which companies are offering the most jobs. We provided a slider for users to adjust how many companies they want to compare at a glance as well.

This chart allows users to see which companies offer the most jobs and thus apply to them. Furthermore, seeing multiple companies from the same industry on the visualisation could help users identify growth industries.

We discovered through the visualisation that A\*STAR Research Entities was the top company in terms of job openings. They are followed by NUS and NTU. This showed us that government affiliated companies seem to have a higher posting rate on the platform. However, this could also be due to the fact that our portal is a government one, and thus these government affiliated companies could have been strongly encouraged to contribute postings on this portal. Meanwhile other companies, big or small, could be posting on other sites, thus our chart cannot give a full picture of the whole country.

## **Salary vs University**



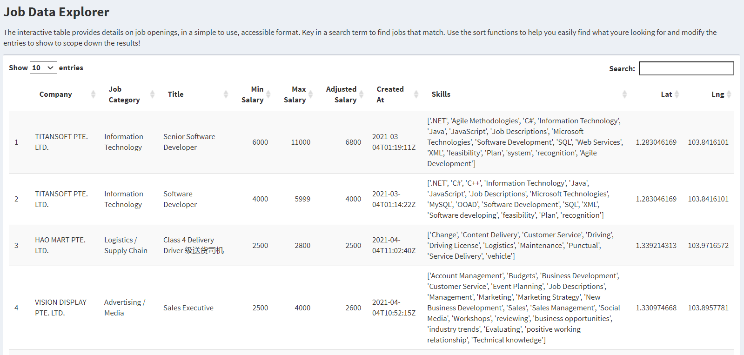
For this visualization, the data set we are using comes from the Graduate Employment Survey for Universities Graduates. This data is obtained directly from data.gov.sg and it's only limited to the 2018 data. To do this, we decided to use a bar chart and sort the various universities’ graduate salaries in ascending order to show where the chosen degree is placed amongst the salaries the graduates of the other universities are getting.

The visualization is done by firstly grouping by the degree and the university followed by getting a mean of the gross **MEDIAN** monthly salary. The median is a fairer view of the salary as compared to the **MEAN** because the mean can often be skewed by anomalies. To make the visualization tighter, we changed the names of the universities to its abbreviation so that the text does not overlap with each other.

The filter is done by getting the bin of the degrees as well as the bin of the universities such that all degrees available in the data can be selected. We then create a separate data frame for the degree itself so that we are able to get the value of the gross median monthly salary of the degree itself. After which, this data is then appended as “Chosen Degree” to the above data frame so that it can all be shown in the visualization.

This helps the graduate to know what salary to ask for when the employer asks for it. Moreover, it also gives the user a look at how the others are getting paid.

## **Data Exploration – Jobs**

****

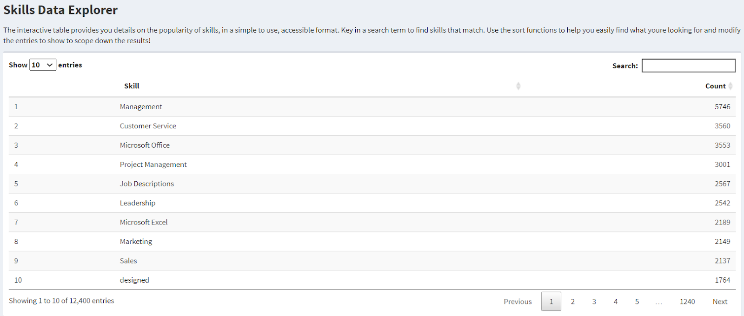
On top of all the job visualisations above, we also provided an interactive table to allow the user to view the job postings in a tabular manner. This provides the user an alternative but more customisable way to view and compare the job postings as the table allows for sorting and searching features.

To create the interactive table for the jobs, we use the dataset we scraped from the MyCareersFuture.gov website. Then, we used data.table and DT libraries and simply added the jobs CSV into the parameter to generate the table.

This interactive table includes details such as the company, job category (industry), title, salary ranges, date of job postings, etc. The user can click on the various columns to sort the table by that column or directly search for a particular job whose details he/she is only interested in. With the sorting feature, users can easily compare the salaries between the different job postings, as well as the companies that are hiring. By searching in the search bar, the user can also indicate a specific value and the table will retrieve the results accordingly.

Overall, this table provides the user an alternative way to view the job postings, compare amongst them and allow fast, customisable search for a particular job or job detail.

## **Data Exploration – Skills**



On top of the Skills vs Job visualisation above, we also provided an interactive table to provide the user with an easier way or rather a tabular format to view the relevant skills the market demands or search for the relevancy or popularity of a particular skill he is interested in. This would be useful if the user wants to find out whether the skills he/she is learning right now is relevant in today's job market. It allows for easier customisation for the user too.

To create the interactive table, we had to first segment out the skills from the skills column in the dataset we scraped from the MyCareersFuture.gov website. Then, we group the skills based on how many times each skill appeared in the job requirements. Afterwards, we exported it into a new CSV called Skills.csv which contained every unique skill name that appeared and its respective count number which is the number of times the skill appeared throughout the jobs. Then to create the interactive table, similar to the above interactive table on jobs, we utilised data.table and DT libraries and simply added the CSV into the parameter to generate the table.

The user can click on the columns to either sort the table via skills’ alphabetical order or count order which would translate to their respective popularities and relevancies. The user can also directly search for a particular skill if he/she is only interested in that particular skill.

Ultimately, this table provides an alternative way for users to compare the skills and more customised in a sense where they can directly type into the search bar too.

# RECOMENDATIONS AND SOLUTIONS

We recommend that job seekers utilize the skill demand feature to prioritize the hard skills that are more sought after to increase their employability. This could also be used to know which words to fill in the resume to increase the chances of your resume being picked from the pile of resume submissions.

We saw some abnormalities in the data where March’s count of job listings were much higher than the rest of the months. However, this coincidentally aligns with the timeline where fresh graduates are usually on the lookout for jobs. The users of the application can use this to potentially keep track and observe the seasonality trends of the job postings over time.

This application also has another use case where employers could find actionable insights to the hiring market by understanding what their competitors are listing. They would be able to know what’s the average wage like for the industry, whether they are on the right platform and when they should begin their hiring process so secure the talents before everyone else starts doing so.

# LIMITATIONS / CHALLENGES

Standardizing data sets across platforms is a challenge. With such limitation we can only observe one platform currently. However, we could increase outreach to Glints, JobsDB, SkillsFuture. The system is scalable and ready to integrate datasets

Frequent and consistent scraping is needed as Jobs position taken may result in delisting. This results in an inaccuracy where not all 100% of all jobs are captured. Also, total number of vacancies are not explicitly publicised for each job available.

We are only investigating Fulltime and Permanent job offers as Internship and Part-Time offers are not in the picture due to large discrepancies and incomplete data.

Theres also limitation where we are Unable to find relationships between other datasets (E.g. GES and Job Listings) It required additional cleaning and analysis to draw clearer relations as well as the manual cleaning would be very time consuming.

CONCLUSION

We plan to integrate datasets from additional data sources such as LinkedIn, Glassdoor and Indeed. By having the ability to plot platforms against each other, we can better understand the various niches and focus of the platform. We would also be able to know where the higher paying jobs are posted and also which platforms are more popular amongst the employers, which platforms are updated regularly.

Moreover, we plan to expand on our GES analysis by matching the degree name to various job categories. The concept of doing this involves creating “sets of job categories” that belong to the degree such as: “Accounting”: {“Auditing”, “Corporate Finance”, “Business Administration”}. We would also clean up the skills such that we remove the anomalies where employers key in office but in fact they meant Microsoft Office. This would consolidate and centralise all skills to only have one representative header. We could potentially leverage on the improved data to predict the salary the graduate would receive by mapping a list of skill sets to each degree. If the user possesses a rare skill, it could potentially be used as a “multiplier” for the user’s salary.

Finally, we would also want to work with government agencies to help them explore potential skills to prioritize in the curriculum as well as understand the job trends in the job market. We could be an independent contractor that was retrieving important information across the market. By creating a set of standard operating procedures to follow, this could potentially create a new data standard in the market throughout all job agencies and hiring companies. There would be agencies and companies who would subscribe to our services to actively and continuously provide them with timely data about the market status so that they will be aware of market shifts almost immediately.

##### References

MyCareersFuture Singapore. (n.d.). Retrieved April 01, 2021, from https://www.mycareersfuture.gov.sg/

We suggest that you use a text box to insert a graphic (which is ideally a 300 dpi TIFF or EPS file, with all fonts embedded) because, in an MSW document, this method is somewhat more stable than directly inserting a picture.

To have non-visible rules on your frame, use the MSWord “Format” pull-down menu, select Text Box > Colors and Lines to choose No Fill and No Line.