**Web3 Development Trend and Advice on Job Seeking:**

**Based on Data Mining of Job Postings on a Web3 Career Site**

Course Code: COM 5507

Course Title: Social Media Data Acquisition and Processing (2022/2023)

Course Instructor: Dr. Xiaofun LIU

Group-6

ZHOU Xingyu 57274685 ZHOU Hongxin 57452863

PENG Jinfeng 57452550 LUO Lihuan 57525301

Communication and New Media

City University of Hong Kong

Content

[1 Introduction 3](#_Toc121323797)

[2 Results 3](#_Toc121323798)

[2.1 Web3 job posting basic information 3](#_Toc121323799)

[2.2 A subdivision of the demand and salary status of various types of web3 jobs 4](#_Toc121323800)

[2.3 The current hot topics and development status of the web3 job market 7](#_Toc121323801)

[2.4 .Remote-working trend&global geographical distribution 10](#_Toc121323802)

[3 Methods 13](#_Toc121323803)

[3.1 How are the data collected 13](#_Toc121323804)

[3.2 How are the data cleaned 19](#_Toc121323805)

[3.3 How are the data processed 22](#_Toc121323806)

[4 Visualization and results 24](#_Toc121323807)

[5 Conclusions 24](#_Toc121323808)

[Reference 25](#_Toc121323809)

[Appendices 25](#_Toc121323810)

## Abstract

Based on data mining from a job site endorsed by industry insiders, this study is dedicated to perceiving the development dynamics of the industry as a whole by cutting through the microscopic perspective. We found that the industry is currently at the beginning stage of healthy development and is in great need of technical talents and non-technical talents who are good at top-level design and proficient in finance, cryptocurrency, NFT, blockchain, and marketing, while a large number of internship positions exist and are friendly to candidates with no experience. Meanwhile, this study also finds that the trend of decentralization is emerging, but North America is still in the lead with its technological advantages, and East Asia, especially Hong Kong SAR and mainland China, has greater potential for development with its huge population and consumer market.

## Individual Contributions

ZHOU Xingyu（57274685）:

Leader, Scrape the basic information and detail page text

ZHOU Hongxin（57452863）:

Scrape the basic information and detail page text

PENG Jinfeng (57452550):

Scrape the basic information and detail page text

LUOLihuan (57525301):

Visualization and Analysis

## 1 Introduction

The term web3 refers to the putative next generation of the web’s technical, legal, and payment infrastructure – including blockchain, smart contracts, and cryptocurrencies. For its advocates, the peer-to-peer character of web3 means it represents a more equitable vision for more job opportunities and talented people to communicate. (Gilbert. 2022)[1]

Web3 is leading a wave of the next generation of web services that even many dominated Web2 platforms (Facebook, Amazon, Apple, Google, and other big tech companies) are keen to ride. However, the lack of Web3 background for Web2 developers hinders easy and effective access and transition. On the other hand, Web3 companies are also keen to have access to top talent and human resources from traditional Web2 companies. (Yu et al. 2020)[2]

Our research aims to provide a slice-and-dice analysis of the industry, using a typical case to identify current market needs and predict future development. For Web2 developers who want to start a new career， we can provide a reference for the future development direction. School students can be given inspiration on web3 skills acquisition.

Although many positions are currently remote, through the actual posting of job information we found that there are still many positions required to go offline, so we added the analysis of geographic location information for the future development of the career of job seekers for reference. This can also reveal the level of attention paid to the industry by different regions and provide a grasp of the overall development trend.

## 2 Results

### **2.1 Web3 job posting basic information**

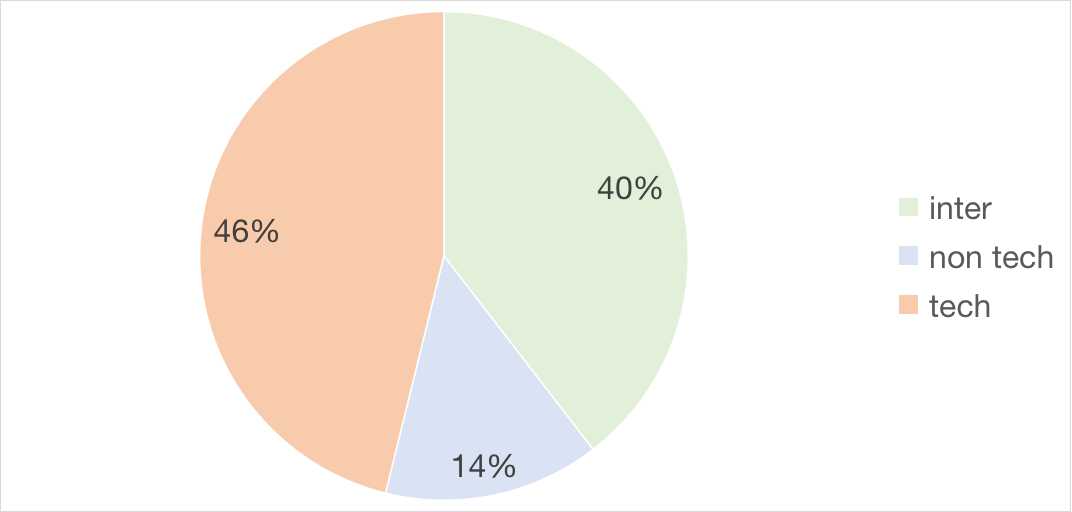
**Job names&fields**

After data cleaning, a total of 3072 valid messages were collected.

|  |  |  |  |
| --- | --- | --- | --- |
| overall | inter | non tech | tech |
| 3072 | 1215 | 439 | 1418 |

Table 2.1 job names&fields

46% were technical positions, 40% were internship positions, and only 16% were non-technical positions.



Graph 2.1 job names&fields

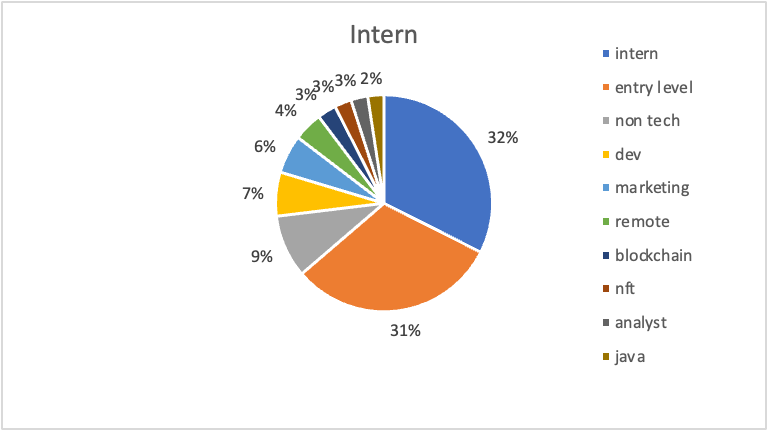
This is a reflection of the fact that the web3 industry is in the development stage, still with the basic platform building, back-end programs, and code writing as the main job search needs, and is trying to attract fresh blood to explore the role of interns

We grasp the job classification tag information, which will be in the next part of the popular top ten analysis.

### 2.2 A subdivision of the demand and salary status of various types of web3 jobs

We selected the ten positions with the highest popularity and further ranked them among the top ten to present subdivision of the demand status of various types of web3 jobs

**Top 10 demands of internship**



Graph 2.2 top 10 demands of internship

The top 10 demands of internship are entry level, intern (these 2 means uncertain),non tech,dev, marketing,remote (uncertain), blockchain,NFT,analyst and java.

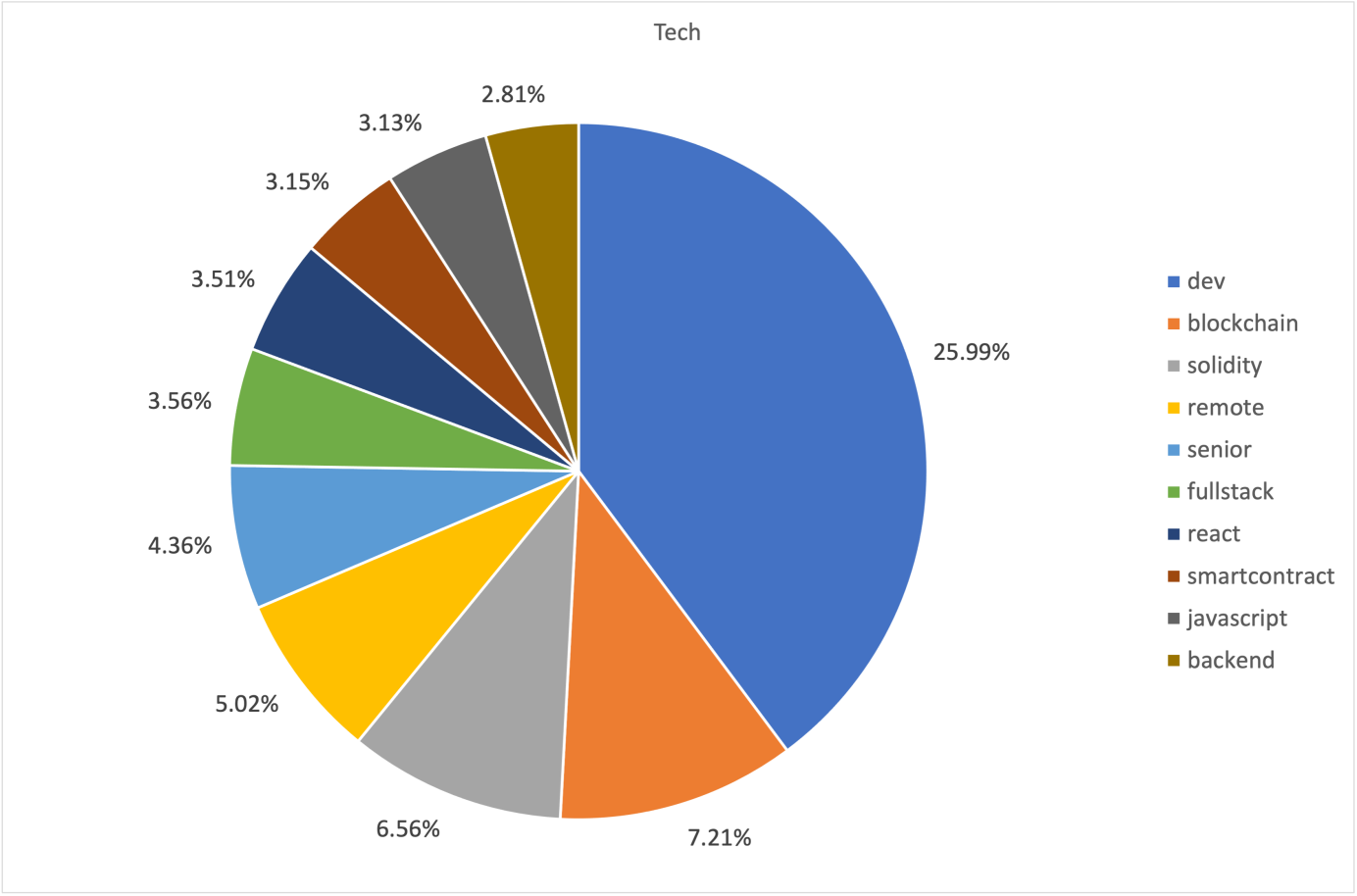
Combined with the current situation of industry development, we can draw the following conclusions.

The web3 industry is in its initial stage, and the skill requirements for interns are currently ambiguous, mostly for the intern.

For non-technical internship positions, there are many opportunities to try, especially marketing. there are also many development positions in technical positions, of which java is a popular language.

Hot topics, such as blockchain and NFT, also have many opportunities for internships

**Top 10 demands of technical**



Graph 2.3 top 10 demands of technical

The top 10 demands of technical positions are dev, blockchain, Solidity, remote, senior, full-stack, react, smart contract, javascript, and backend. It contains several industry hot topics

Firstly, Solidity is a curly-bracket language designed to target the Ethereum Virtual Machine (EVM). It is influenced by C++, Python, and JavaScript. You can find more details about which languages Solidity has been inspired by in the language influences section. And it is an object-oriented, high-level language for implementing smart contracts. Smart contracts are programs that govern the behavior of accounts within the Ethereum state.

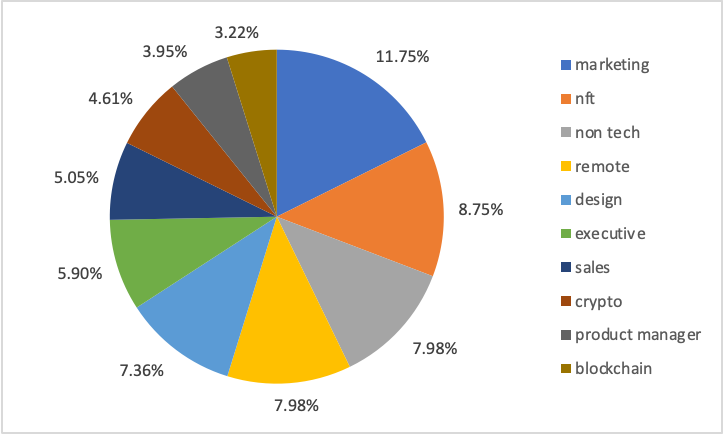
Secondly, the term “full stack” refers to the technologies and skills needed to complete a project, with each individual component being a stack. Stacks can be mobile, Web, or software specific. Typically, a software engineer will focus on one part of development, either the front end or the back end.

Thirdly, React is a JavaScript library created by Facebook, a User Interface (UI) library, and a tool for building UI components.

To sum up, we can find that all kinds of web3 companies are currently looking for very mature technical professionals who need to be familiar with all kinds of programming languages, especially java-based.

Candidates are also required to have the ability to independently complete project planning and implementation, so in the technical aspect, web3 has a large industry barrier.

**Top 10 demands of non-technical**



Graph 2.4 top 10 demands of non-technical

The top 10 demands of non-technical positions are marketing, NFT, non-tech(uncertain), remote, design, executive, sales, crypto, product manager, and blockchain.

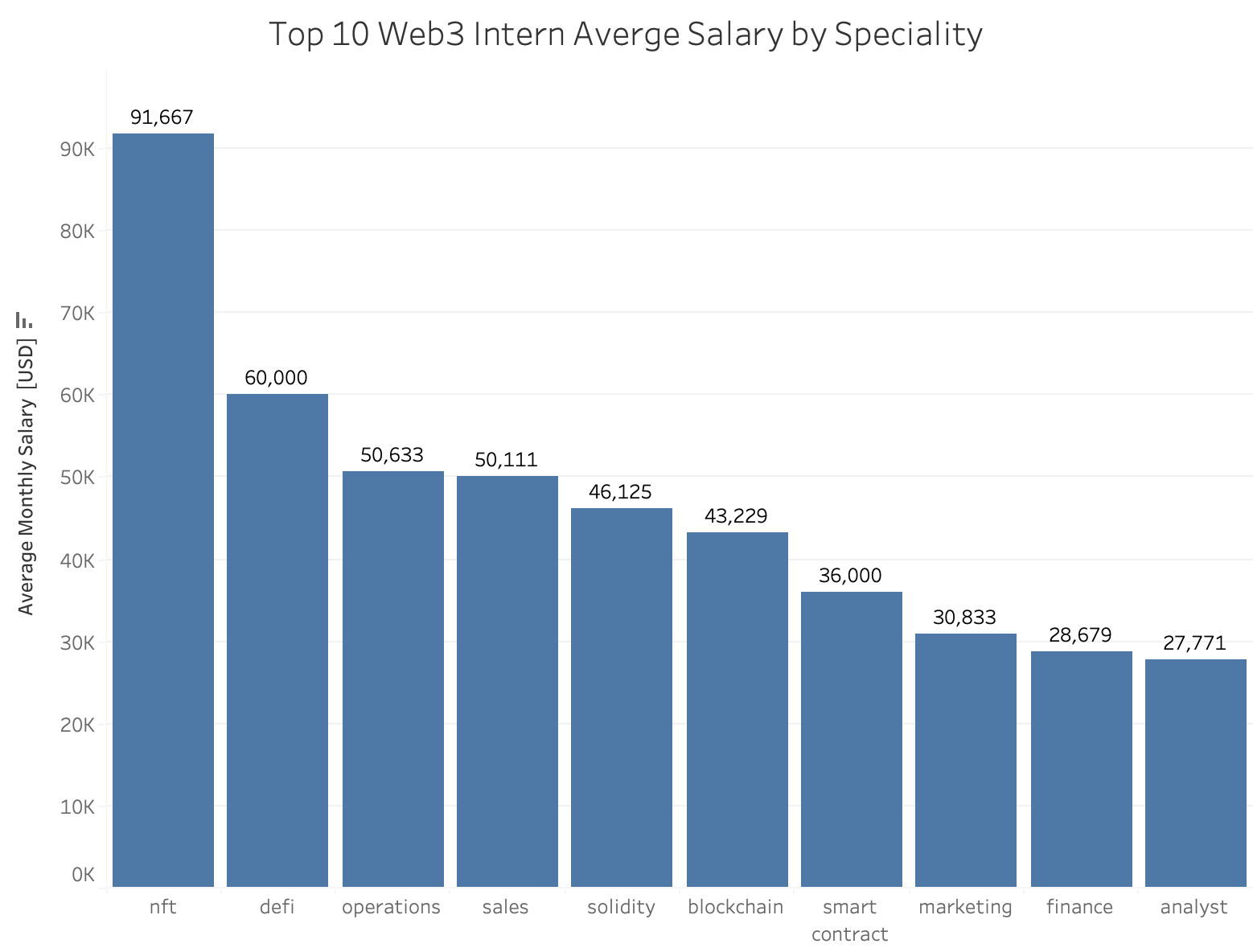
Professor Almeida from Cornell University predicted in 2013 that Web 3.0 will have a significant effect on users and businesses. Today it has already changed how companies use the information to market and sell their products, as well as operate different businesses. (Almeida et al. 2013).[3]

Therefore, in non-technical areas, business-related positions such as marketing, product management, and sales are very popular. At the same time, there are many job opportunities for design positions.

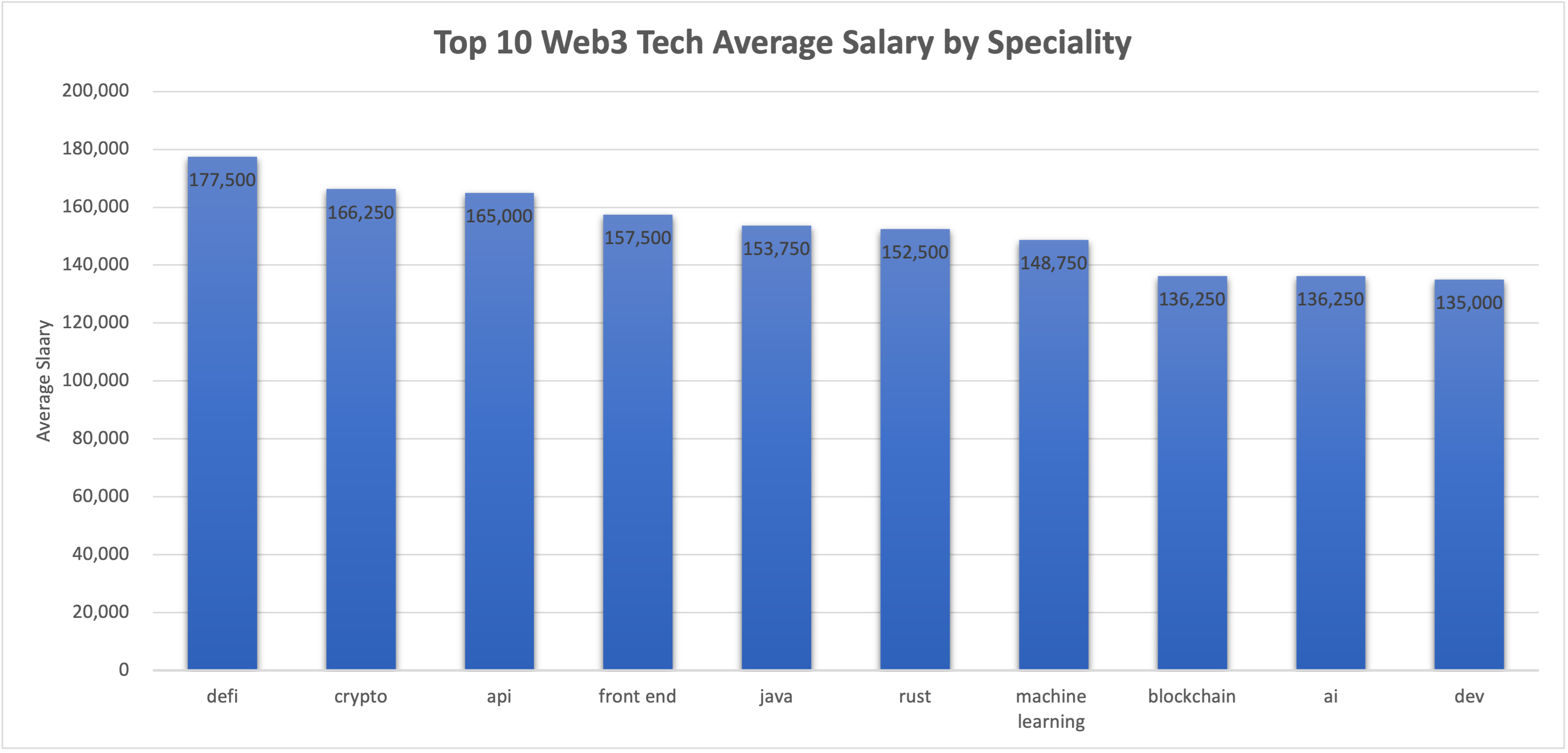
**salary status**

Since the salary provided by this website is an interval, in order to more intuitively show the salary of different positions, the median of the interval was calculated using the function first, and then the average was calculated.

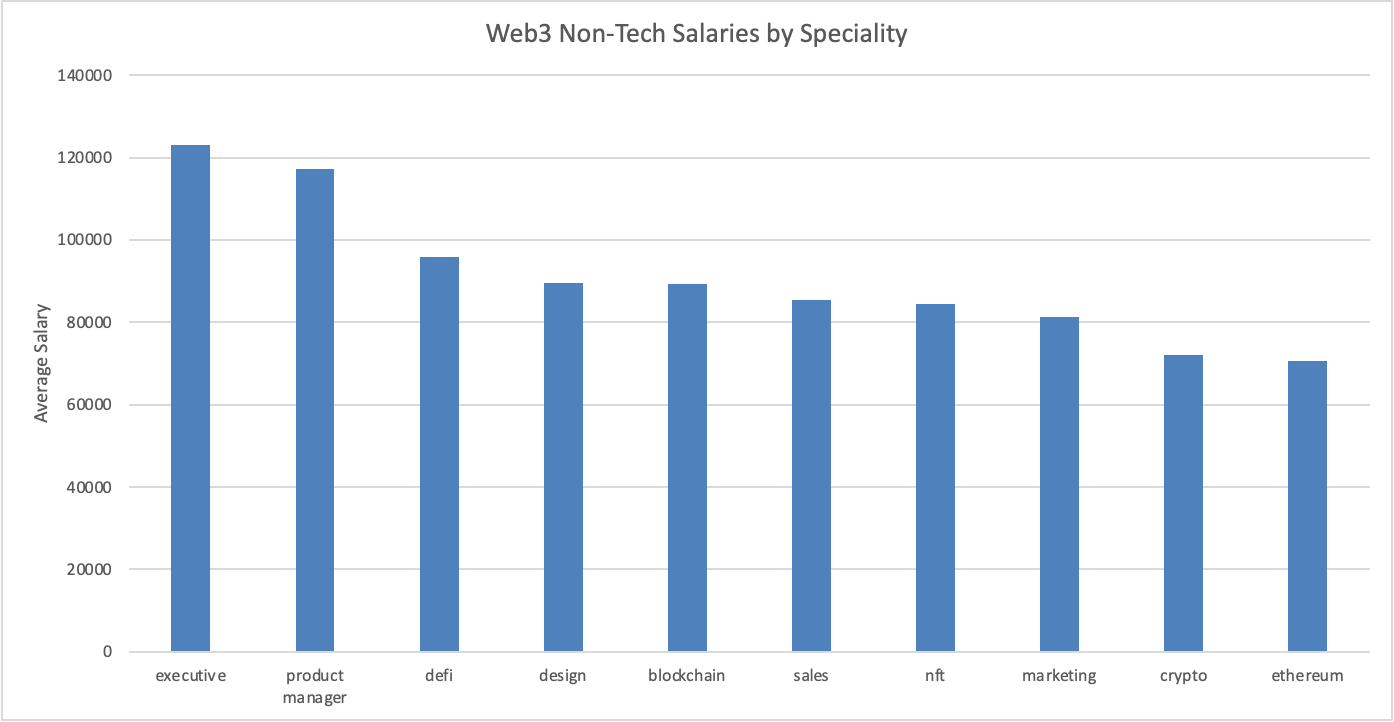
The bar chart shows us that the most popular internship positions with the highest salaries are NFT related positions. Among the most popular technical positions, the highest salary is DeFi (Decentralized Finance), which also reveals the current industry trend.



Graph 2.4.1



Graph 2.4.2

Graph 2.4.3

### 2.3 The current hot topics and development status of the web3 job market

Analysis through the word cloud, we can intuitively see the most popular words on the web3 job market, and use this to infer the current development of the industry and the hot topics that employers care about most.

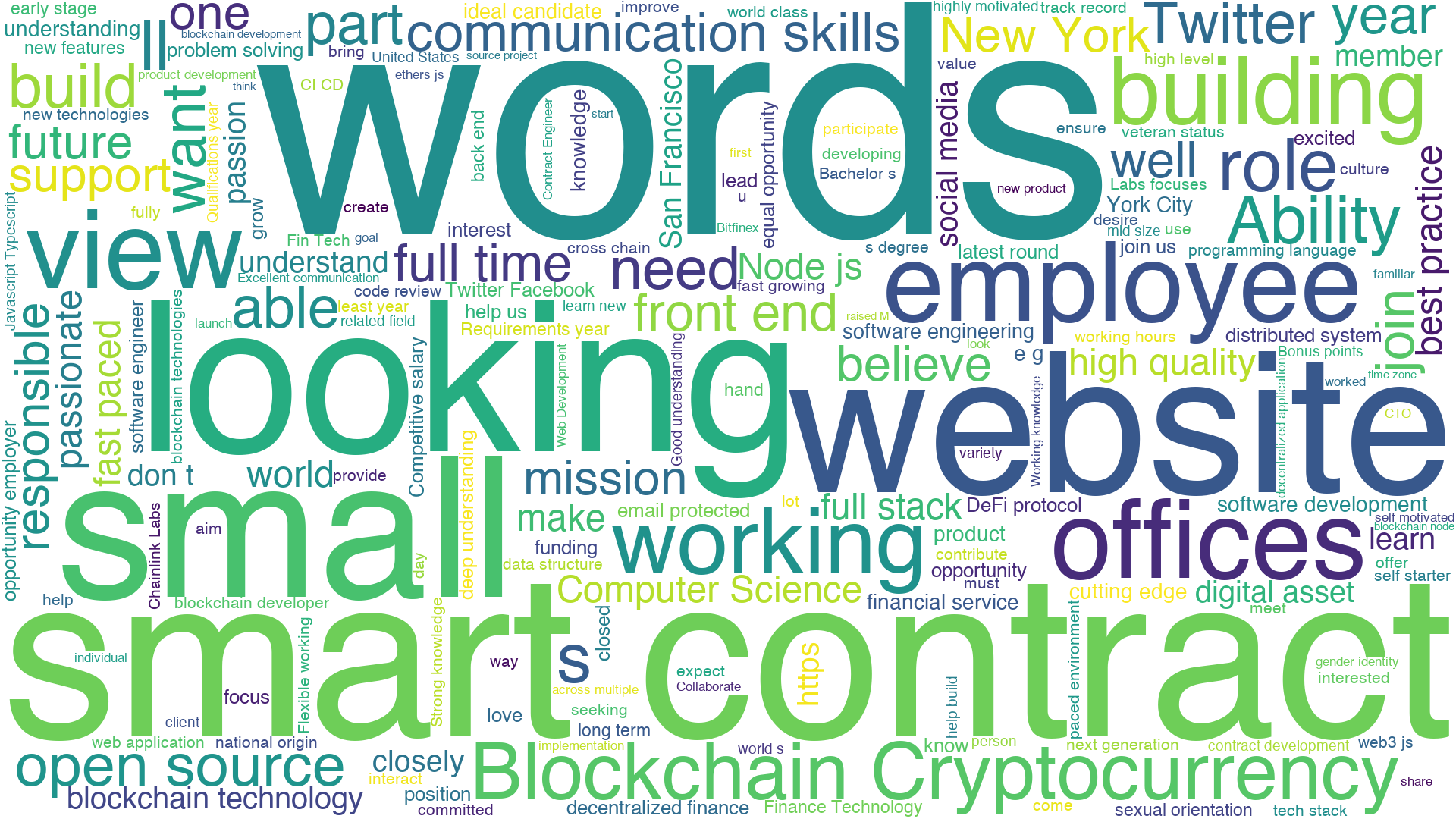
**The current hot topics and development status of internship**



Graph 2.5 Words cloud of internship

From the overall content of the internship position gives a positive feeling, binance and other cryptocurrency companies are very active and are trying to attract more new people.

**The current hot topics and development status of technical**



Graph 2.6 Words cloud of technical

The current hot topics in the technical field as mentioned in the previous section, smart contracts, blockchain, cryptocurrencies are still popular, and there is still a great demand for such basic as website building.

**The current hot topics and development status of non-technical**



Graph 2.7 Words cloud of non-technical

The biggest trend in non-technical areas is product, and other topics that revolve around marketing. Other flash points are more similar to technical positions.

### 2.4 .Remote-working trend&global geographical distribution

Graph 2.8 Working mode classification

Through the analysis in the previous sections, we can see that remote work is common to a number of web3 positions. Especially during the COVID-19 pandemic. Even as the world opened up more this year, remote work is the new normal. According to a survey by Jamstack,83% of developers are working remotely more than half of the time, and 76% say they've maintained or increased their frequency of working remotely in the last year. Remote work flexibility and career growth opportunities were the most frequently cited reasons for leaving a job, followed by compensation. (Jamstack. 2020)[4]

Therefore, in the process of analyzing the workplace, we paid attention not only to the development of each region but also to the trend of remote work.

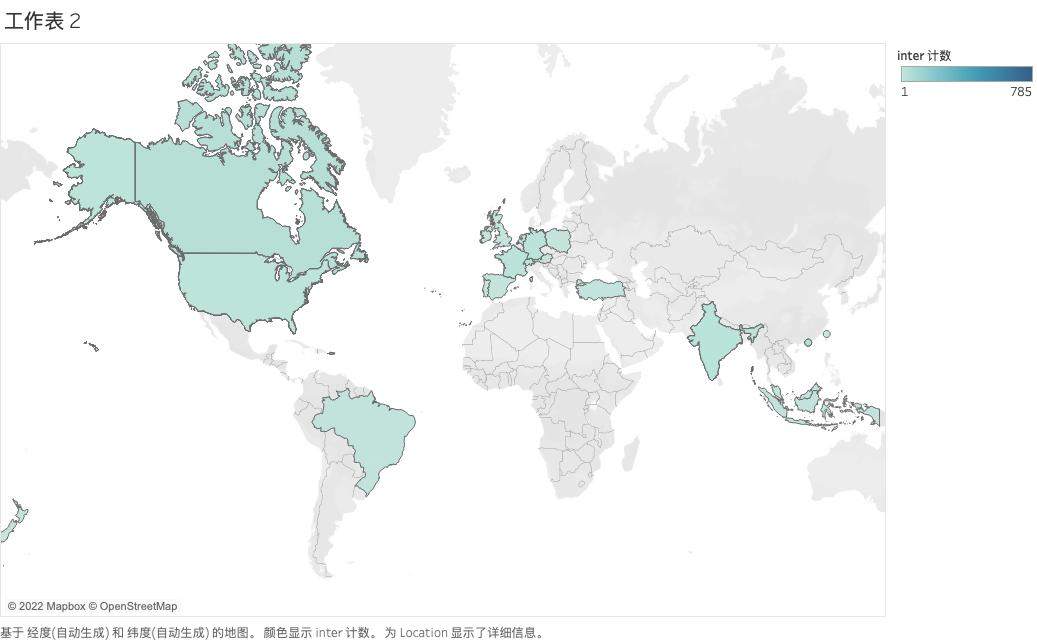
This bar chart shows that the existing job postings on this website clearly indicate that remote work is still rare and does not meet job seekers' expectations for job flexibility.

Many of these internship positions are uncertain whether they are remote or not and should have more negotiation space.

Unlike what we expected, there are not many technical positions that can be performed remotely.

Combined with the word cloud we can find that, unlike web2, web3 technical positions require more communication skills and teamwork abilities, which may present a challenge to the remote working model.

**Internship geographical distribution**

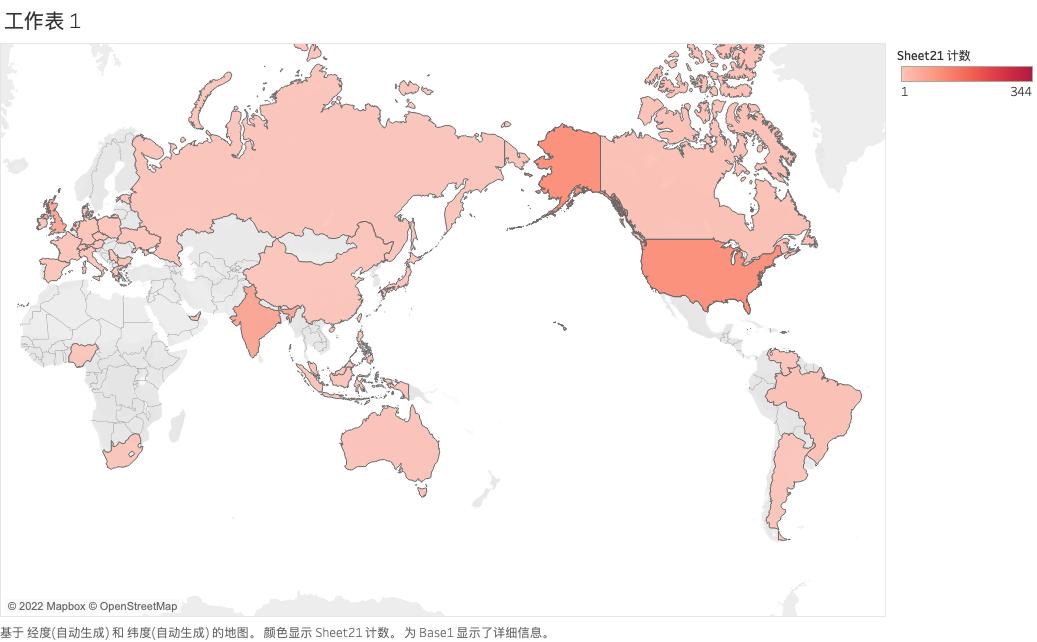


Graph 2.9 internship global geographical distribution

The visualization of the location of the internships shows that there is a certain and even distribution of demand in Europe, America, and Southeast Asia as well as in Hong Kong SAR and Taiwan.

We note that there are many job requirements in South America. Combined with the recent news, Devcon 6（an annual conference for all Ethereum developers, researchers, thinkers, and makers）is held in the South American country of Colombia. Through these, we can find that web3 is in the process of disrupting the traditional Internet industry pattern and is promoting further decentralization.

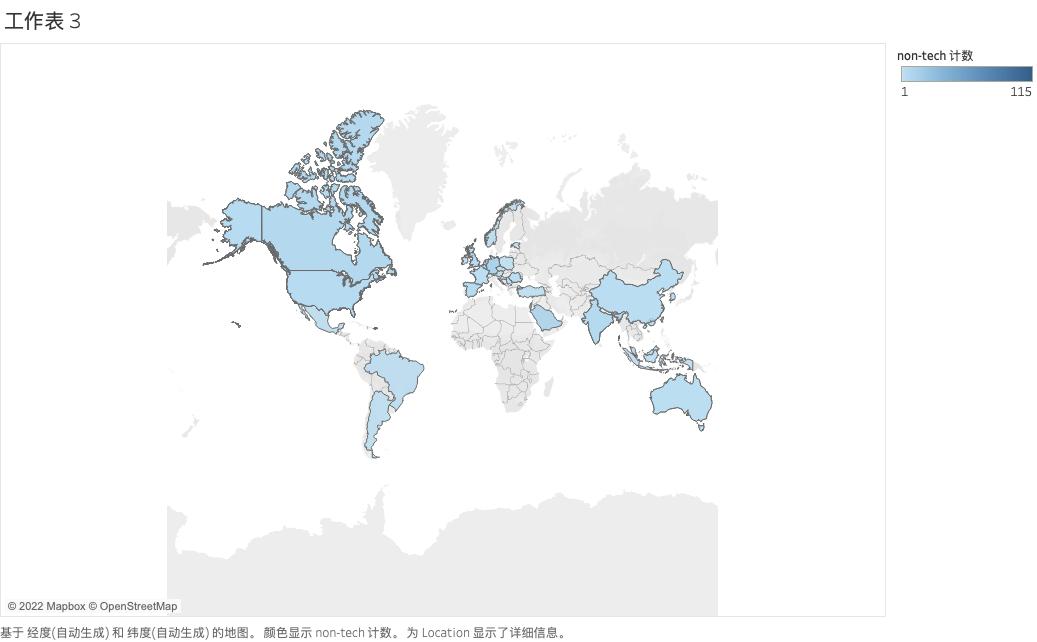
**Technical geographical distribution**



Graph 2.10 technical global geographical distribution

The visualization of the location of the technical position shows that demand for technical jobs is heavily concentrated in North America, which continues to have a first-mover advantage.

**Non-technical geographical distribution**



Graph 2.11 non-technical global geographical distribution

The visualization shows that for non-technical positions, the global demand is more even. Most of the positions in mainland China are in Shanghai. Hong Kong has more opportunities compared to other nearby countries and regions.

## 3 Methods

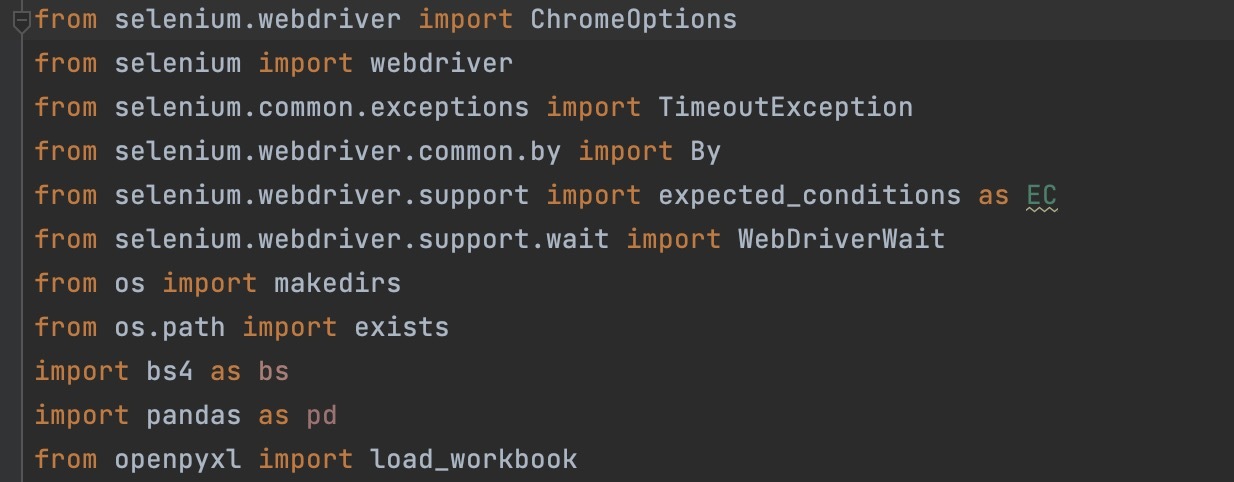
### 3.1 How are the data collected

For list pages, selenium, beautifulsoup, and pandas are the main libraries we use to scrape the data. We first use selenium to click the Next button to browse all the pages, then use beautiful soup and for-loop structure to locate all the elements we need on each page, and finally use pandas to store all the data in a excel file. All the labels we need is shown below.



Graph 3.1

These are all the libraries we use.



Graph 3.2

We create some variables to store some basic information, like website, total page numbers, and the folder we use to put our results.



Graph 3.3

First, we use selenium to open the index page and return its html code. This is the step we use to start our programme.



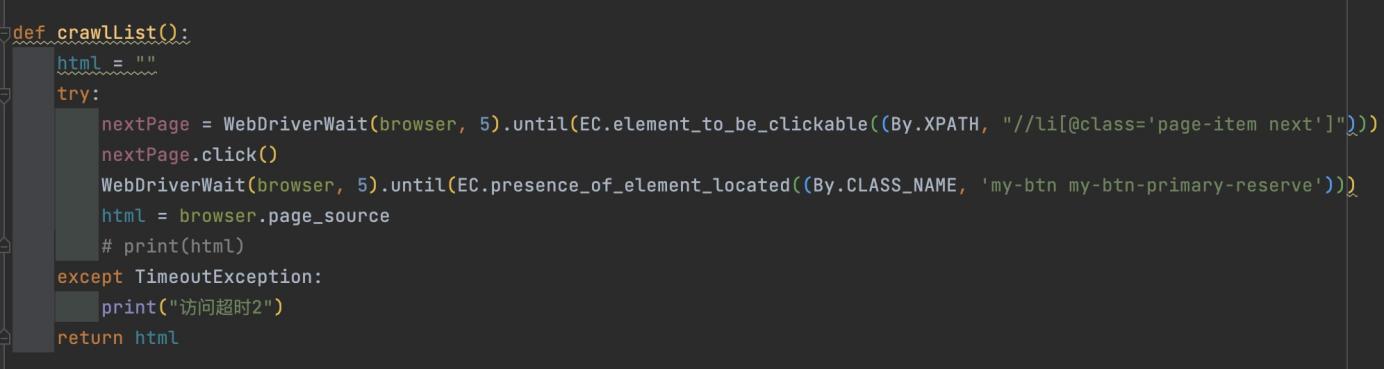
Graph 3.4

Then, we use beautifulsoup to analyze the html code and locate all the labels we need. We extract the text in those labels and put them in different lists. In this step, what we return is a collection of lists.



Graph 3.5

After that, we write some codes to find the next page button so that we can click it after we get all the labels in this list page.



Graph 3.6

We use pandas to store that collection of lists, each list is put in a column of a table. We write the header and export the lists to a excel file.



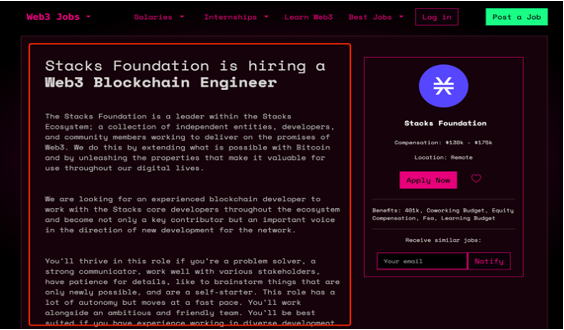
Graph 3.7

Finally, at the end of this python file, we write the body of all the function, by using the for loop, we make the whole programme browsing all the list pages one by one and append the new data on each list page to the original excel file.



Graph 3.8

For the detail pages, the text we need is shown below, when we click the iob names on the list page, we can enter the detail page. So our method is to locate all the elements containing job names on each list page so that we can put all the urls in a table. Then, by visiting all the hyperlinks in the table, we use beautiful soup to locate all the detail text in each hyperlink and put them in an excel file.



Graph 3.9

To get all the detail page urls, we use the same method(selenium) of capturing data from the list page. But when we located the link, we choose to use xpath to analyze the html, @href function is used for extracting all the urls on each list page.

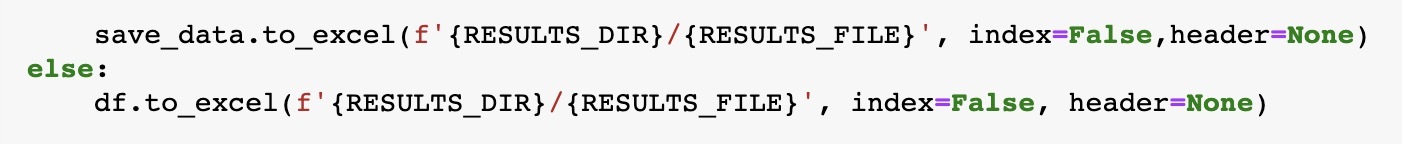
IMG_265

Graph 3.10



Graph 3.11

Then we put all the urls in an excel file, the header is none so that we can traverse the table more easily.



Graph 3.12

But when we inspect the excel file, we found that all the urls lack the domin name, so we use openpyxl to load the table and use pandas to add domin name for each url in the table.

IMG_268

Graph 3.13



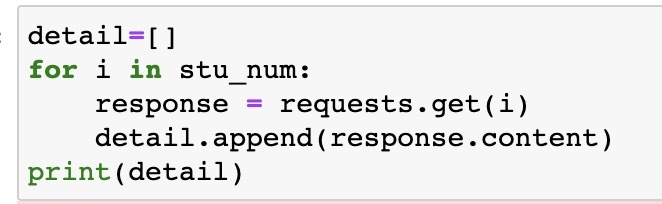
Graph 3.14

After we added all the domin names, we reload the table and put the url column in a list.



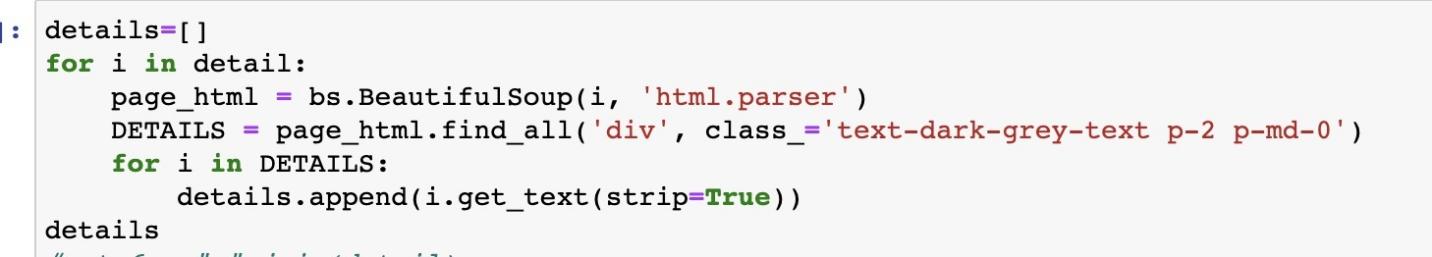
Graph 3.15

We create an empty list. Then we use request to open all the urls and return their html codes. All the codes are put in the list.



Graph 3.16

We use beautifulsoup to analyze all the html codes and locate the element we need. Then we get all the text in those web elements and put them in a list.



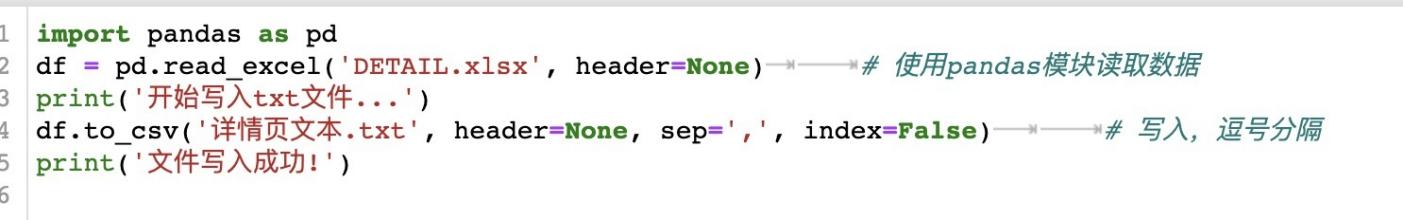
Graph 3.17

Finally, we use that list to create a dataframe and use pandas to export a excel file with out header.



Graph 3.18

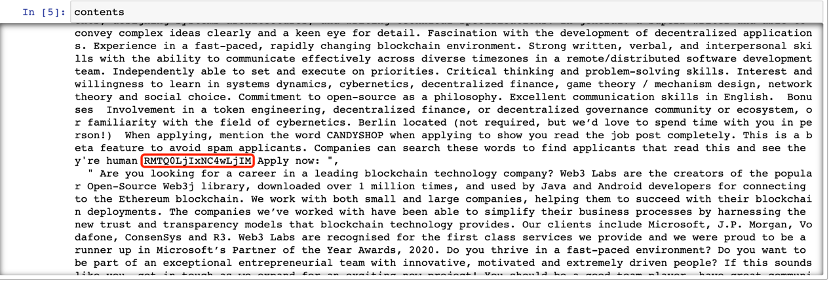
Because we want to use the detail page text to create wordcloud, we transform the excel file to a txt file.



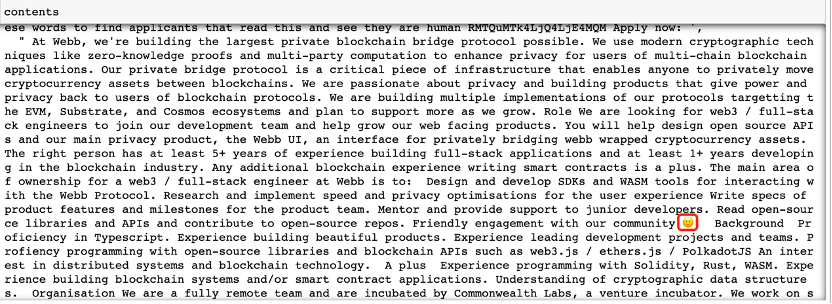
Graph 3.19

### 3.2 How are the data cleaned

After we scraped the detail page text data for each position, we found that the crawled detail page text data contained some strange, garbled characters and emojis, and punctuation marks. We're not sure if this will affect our subsequent data analysis, but to avoid problems, we decided to clean this part of the text data to some extent, removing the gibberish and emoji, and punctuation marks.

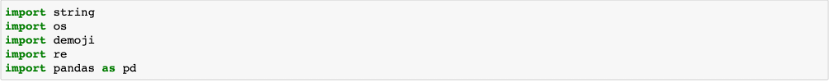


Graph 3.20 Text data of detail page



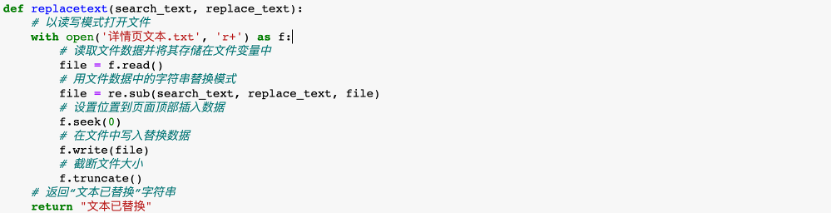
Graph 3.21 Text data of detail page

For this purpose, we need to use “string” module, “os” module, “demoji” module, and “re” module. We import the modules we need in Jupyter notebook.



Graph3.22 modules(code)

We first define a function “replacetext(search\_text, replace\_text)”: opening the details page text data we captured in read-write mode, that is, "details page text.txt", ("r+" means readable and writable. If the file does not exist, an error will be reported. ) storing the data in the “file” variable, using “re.sub” to replace the data (that is, replacing “search\_text” with “replace\_text”), setting the position to the top of the page to insert data, writing the data in the “file”, truncating “file” and returning the "text replaced" string lastly.

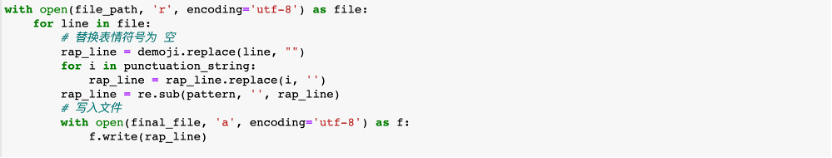
Graph 3.23. define function(code)

We create the "search\_text" variable, which is used to store the text data that we want to replace. Then create the "replace\_text" variable, which is used to store the text data we want to update. Call the "replacetext" function we defined earlier and print the returned statement. After that, create "punctuation\_string" to store the pre-initialized string "string.punctuation" using as a string constant, and use "re.compile" to return the regular expression pattern object. Then remove characters that conform to the “pattern”.



Graph 3.24 data cleaning(code)

Use the “demoji” package to delete the emojis in our scraped text data. At the same time, delete the punctuation marks in our scraped text data and write them to the file.



Graph 3.25 data cleaning(code)

After the above operation, we get the detail page text data of each position after clearing garbled characters, emojis, and punctuation marks.

### 3.3 How are the data processed



Graph 3.26 Raw Data csv file



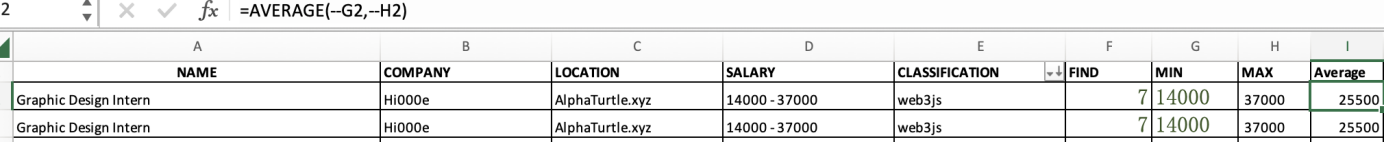
Graph 3.27 Detail Pages txt file



Graph 3.28 Stopwords Cleaning

**Text Processing**

Since we need to do word frequency analysis, we use the stopwords function to clean up our txt files, which would be used for the following word frequency analysis and word cloud analysis.



Graph 3.30

**Numeric values processing**

For Average Salary, we remove "$" and convert "k" into 000. Then get the MAX and MIN salary from the SALARY index. Then we calculate the average salary for each position by using the AVERAGE function.

## 4 Visualization and results

For the basic data organization, we used Excel and created bar and pie charts, for more advanced visualization, we used Tableau for regional heat map distribution.

**4.1 Web3 development current problems and future dilemma**

First of all, the development of web3 is uneven across the globe. Although it practices the principle of decentralization. However, according to the visualization of geographical information, excluding some technical positions, there are basically no relevant positions in Africa. Moreover, cryptocurrency trading platforms such as Binance, Crypto, and Coinbase provide the most demand for Web3 positions. These centralized companies are not only hiring non-technical marketing staff but also recruiting web3 technicians to develop their own web3 projects. It is possible for them to establish the next round of technology monopolies as they expand on the global market. Since the recent collapse of FTX, however, centralized exchanges have been viewed as unreliable.

Second, remote work has become an irreversible trend, but the cost of communication and collaboration is still high. web3 needs a more ground-breaking technological revolution to enhance the possibility of remote work, thus improving the productivity of the entire industry and truly connecting with the metaverse and other related industries in the future.

## 5 Conclusions

First of all, the web3 industry is in the development stage. Cryptocurrency, blockchain, NFT, smart contracts, and other finance-related topics are still the hottest topics.

Second, the development of web3 is uneven across the world, Decentralization is on the way but first-movers have clear advantages. For example, according to the visualization of geographical information, excluding some technical positions, there are no relevant positions in Africa. North America, particularly Silicon Valley, where there is a good web2 infrastructure, is showing great attractiveness for talent at the technical level, and Southeast Asia and China, with their huge populations and markets, are bringing together more non-technical business and supply chain jobs.

Second, Web3 career fields have become characterized by remote work due to the nature of Internet technology and regional heterogeneity. On the one hand, Web3 is democratically accessible to users around the world who are online. On the other hand, Web3 technology still has class or technical limitations. As the infrastructure technology of Web3 is based on computer science and Web2 technology, in Web2 developed regions, such as Silicon Valley in the United States and Shenzhen in China, a vast pool of technical talent provides the impetus for rapid development of Web3. As in other Web2 developing regions, Web3 requires the recruitment of talent from around the world, creating a strong demand for remote workers. Generally, Web3 needs a more ground-breaking technological revolution to enhance the possibility of remote work, thus improving the productivity of the entire industry and truly connecting with the metaverse and other related industries in the future.

But for newcomers who have never entered web3, there are still many opportunities.

In summary, according to our data, the industry is growing but presents a large lack of talent. And in the future, how to improve the efficiency of remote working mode is also a dilemma

## Reference

[1] Gilbert, S. (2022). Crypto, web3, and the Metaverse.

[2] Yu, G., Wang, X., Wang, Q., Bi, T., Dong, Y., Liu, R. P., ... & Reeves, A. (2022). Towards Web3 Applications: Easing the Access and Transition. arXiv preprint arXiv:2210.05903.

[3] Almeida, F., Santos, J. D., & Monteiro, J. A. (2014). e-commerce business models in the context of web3. 0 paradigm. arXiv preprint arXiv:1401.6102.

[4] Jamstack. “Jamstack Community Survey Results（2022）Jamstack.” Jamstack.org, 7 Nov. 2022, jamstack.org/survey/2022/. Accessed 6 Dec. 2022.

## Appendices

7.1.Table＆Figure(s)

7.2.Python code with detailed notation

7.3.Original dataset (e-copy only)