A Comparative Analysis of Global Remote Work Salaries

Yutong Wen
yutong.green@outlook.com
GitHub

Raseswari Mondal Supti
raseswarimondalsupti@gmail.com

Abstract

Due to the Covid-19 pandemic, remote working has already been a "regular" scenario and more people are being concerned about the salary from working remotely. The report summarizes visual information and analysis results related to the remote salaries. The purpose of this report is to study regional differences in remote ratio and salaries. We have used choropleth map to investigate the remote working ratio in the context of number of jobs and have done hypothesis test to find out whether there is any significant difference in the average salary.We hope our findings from this report can help the job seekers make better decisions.

1 Introduction

The global COVID-19 pandemic, with the emergence of several new and more infectious variants until now, has its drastic effect on the practice of work place as well as workers work patterns within organizations globally. And remote working being used to refer to the term telecommuting in its earliest time [1], has become one of the much needed ways to ensure the productivity from the workers and keep the wheel of economy moving during this pandemic. To get a sense of the effect of remote working ratio on salaries, we have used salary data from the website called FreshRemote.work and for visualization part we used "naturalearth_lowres" from GeoPandas along with the salary data. To draw the global scenario of the ratio of remote working of jobs, we have used choropleth map. We have found a great deal of variation across different countries in the number of jobs following 100% remote working. Through the analysis, we have also found significant differences in the salary between USA and non-USA countries. Our intention behind studying this data set was to provide better information on the salary of remote working and remote work itself on a global scale so that each kind of job seeker (newbies or experienced pros) can make better-informed decisions by using this information.

2 Methodology

Datasets
Our first data set referred as "salary dataset" comes from the FreshRemote.work. It accommodates salary information like different jobs related to different years, locations, remote ratio, experience levels, types of employment, and company size. The working years ranged from 2020-2022 and are divided into two types: deterministic and estimated. In this report, we mainly use data from years 2020 and 2021. Unlike the common definition, the remote ratio in this data set is a discrete number where 0 indicates less than 20% remote work, 50 indicates partial remote work and 100 means more than 80% remote work. Among the four-level of experiences in this data set, and we only studied two: junior level and intermediate level. The data set is published under **CC0** and gets updated monthly and is completely free with the open access to the public.

The second dataset referred as "world dataset", comes from GeoPandas named "naturalearth_lowres". This dataset provides geospatial data for all the counties in the world under polygons or multi-polygons. In this report, we use this dataset as an auxiliary dataset to help display the world map.

Data Preprocessing

Like many other data, the main downside of these two datasets is **incomplete**, indicating the presence of missing values. To start with, the world dataset is excluded of countries as significant as Singapore. Therefore, we have removed data for the missing countries in salary data to match the world data. In the salary dataset, the inclusion of different jobs varies every year, for example, jobs like 'Python Developer' appeared in 2021 but not in 2020. Moreover, there is also discrepancy in the experience level of some jobs. For instance, there is no entry-level position for Principal DevOps Engineer. For meaningful result, we only used the definitive data for 2020 to create the choropleth map, because the data is relatively complete compared to other years. For the salary study, we only focused on the job that occurred both in 2020 and 2021. Another drawback of the salary dataset is **redundancy**, which means the same job may have different recording style in different years e.g. there are two different titles 'Cybersecurity Engineer' and 'Cyber Security Engineer'. We unified the names of the jobs for the same work with different title. There is also some labeling **errors**, for example, France has ISO code -99 in the world dataset. We changed the ISO3 code for United States Minor Outlying Islands and American Samoa to 'USA'. Then we converted ISO3 code from the countries like France into the real ISO3 code in the world dataset. For a better match, we converted the ISO2 code in the salary dataset to ISO3.

3 Visualization

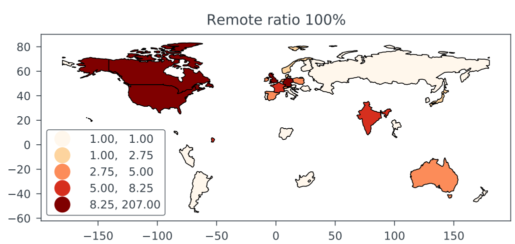
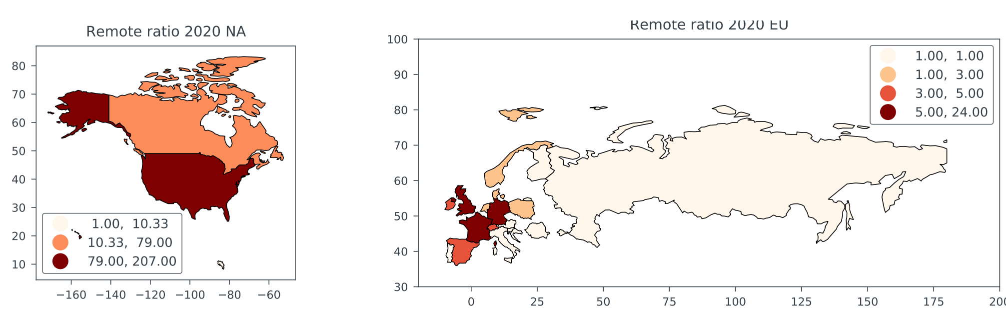


Figure 1: Jobs with 100% remote ratio

The figure 1 portrays an overview of the total number of jobs among different countries globally using the 100% remote working ratio. From the legend we can see that, this map shows countries with at least one job with 100% remote ratio and exclude countries that do not fall into this context e.g. China. Maps using 50% and 0% remote ratio were also excluded because of the severity of missing data. Since the range of the number of jobs is very large and vary significantly between North American (NA) and Europe (EU), in order to dig deeper we plot the next two plots focusing on them.



(a) (b)

Figure 2: North American and Europe

We generate two more plots for NA and EU showed in figure 2 excluding the part generated by missing data. From the NA plot it is visible that USA has the highest number of jobs with 100% remote working whereas Costa Rica shows only one job with complete remote working. In EU, United kingdom has the highest number of jobs for 100% remote working where Germany follows the second position. Italy and Russia shows the least number of jobs for complete remote working.

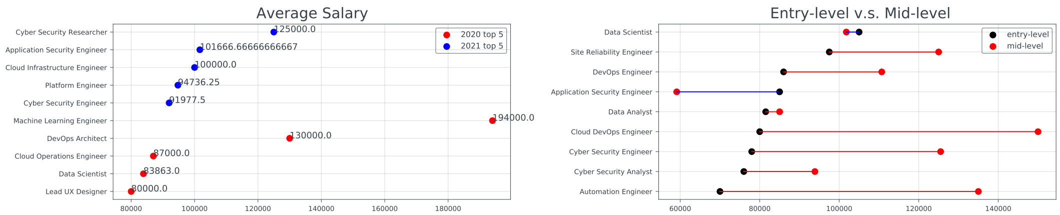
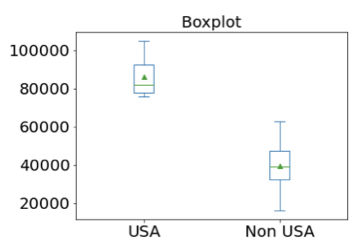


Figure 3: Average salaries of top 5 jobs and a a comparison of entry level and mid level jobs

After looking into the remote ratio, we shift our focus on the salary. As shown in the sub-figure to the left of the figure 3, we compare the five highest paid remote jobs in 2021 and 2020. To conduct that, we calculated the average entry-level salary for each full-time job. The red and blue dots indicate the job in 2020 and 2021 respectively. Clearly, salaries are sparse in 2020 suggesting a large variance. On the contrary, salaries in 2021 are clustered. Following the Choropleth map above, we then looked at the jobs in the United States, because the map suggests that United States has more variety of jobs. In order to explore the fact about change in salary after a promotion in position, we show an illustration of the comparison of entry-level and mid-level salaries for full-time jobs in the United States in 2020. We can see that most jobs pay higher when employees have a deeper level of experience. But there are two anomalies: application security engineers and data scientists. After checking the forms of data, we deduced to the assumption that either our records were wrong or that the employees were simply not U.S. residents. Given the page number constraint, we discuss about this in the limitation suction under fairness issues.

4 Analysis

In this section we conduct a hypothesis test to compare the mean salary differences between the USA and Non-USA region. To achieve that, first we have sampled jobs with same title but different salary from the two regions as shown in the figure 4. Next, we applied **Paired Samples t-test** considering the significance level $\alpha = 5\%$, and the hypotheses are defined as follows:



Job title(size)	**USA**	**Non USA**
Cloud DevOps Engineer(L)	80000.0	32000.0
Cyber Security Analyst(L)	75500.0	15982.0
Cyber Security Engineer(L)	77320.0	44892.0
DevOps Engineer(L)	84000.0	48229.7
Site Reliability Engineer(L)	95000.0	33348.0
Data Scientist(S)	105000.0	62726.0

Figure 4: Boxplot of the salaries of USA and non USA jobs

Null hypothesis: The mean salary differences for same jobs between USA and Non-USA countries are 0.

Alternative hypothesis: The mean salary differences for same jobs between USA and Non-USA countries are not 0.

Before applying the paired samples t-test, we checked whether our sampled data satisfies the assumptions of the test. Our salaries are continuous and our observations are independent of one another. The boxplot in 4 shows that there is no significant outliers in the dependent variables. Then we conduct a normality check because the result of paired samples t-test can only be trusted when the difference is normally distributed. We performed a Shapiro-Wilk test where the null hypothesis is that samples came from a normally distributed population. According to our p values (0.29 and 0.96) from the Shapiro-Wilk test, it had been concluded that the assumption of normality is not violated.

Then we calculate the test statistic and p-value of the paired samples t-test, which are 9.434 and 0.0002 respectively. The result is in favor of our alternative hypothesis that the average salary differs significantly for the same jobs between USA and Non-USA countries.

5 Conclusion and Limitation

5.1 Conclusion

To conclude, it was challenging to develop meaningful outputs as our data comes with its limitation of having lots of missing values. Despite that, we were able to draw the picture of varying numbers for full remote working for specific countries. Also, it was evident that the average salary differed for same jobs between USA and Non-USA countries. On a different remark, from the visualization we can comment a bit ambitiously that, for lucrative salaries young professional can consider working as machine learning engineer or cloud development engineer.

5.2 Limitation

Data Incomplete: The main limitation of this project was incomplete data for different variable such as job title, experience level; especially for 2021. As a result of which our main focus was using data from 2020.

Redundancy: There exists redundancy for variables like employment type with unknown reason as shown in the table below: After cleaning the data we were left with only 6 jobs title in the entry level

Type	Year	Level	Salary	Employee residence	Remote Ratio	Size	Location
FT	2020	SE	128000	US	100	L	US
FT	2020	SE	120000	US	100	L	US

Table 1

as shown in the analysis section. And to deal with the aforementioned situation of redundancy, we took the average of the salary which might not be a correct action but was necessary. Because if we had removed all the redundancy data, we would have left with only two job titles which might have affected our results.

Fairness problem: As shown in the table below, employee with different residence get different payments for the same job under same condition. It could be an error, but it can also consider as a fairness problem. Since, we do not know how the data were handled, we are not extending discussion on this issue but rather mentioning since it might have important effects on analysis using this data.
Correlation: Except for the salary feature, all the other variables were recorded as discrete even

Jobs (company size S)	level	salary	employee residence	remote ratio
Application Security Engineer	Entry	85000	US	100
Application Security Engineer	Mid	8223	IN	100

Table 2

the remote ratio. As a result of which we could not study the correlation matrix to observe which attributes affect the salary and how. It is of importance because it is pretty much evident that attributes like employment type, company size and experience level are correlated to the salary.

References

[1] Jack M. Nilles. Telecommunications and organizational decentralization. *IEEE Trans. Commun.*, 23(10):1142–1147, 1975.