Trends in Remote Work during the COVID-19 Pandemic

Austen Borg
borg22a@mtholyoke.edu

Sophia Hoenig
hoeni22s@mtholyoke.edu

Abstract

This paper reports on the frequency of job postings for remote work over the span of a year from September of 2020 to September of 2021, during the height of the pandemic's effect on workplaces and the job market. We trained a model to make predictions about future trends, with a goal of gaining insight on the more permanent effects the pandemic has had on the prevalence of remote work opportunities, and another model to assess the "remotability" of jobs based on features of their job listings. We find that there is a slight overall decrease in the proportion of available jobs that are remote, as well as three peaks in remote listings in October 2020, March 2021, and August 2021.

1 Introduction

One of the most significant and wide-reaching impacts on day-to-day life as a result of the COVID-19 pandemic has been the shift from in-person to remote work. Many jobs that are currently remote were previously thought to be necessarily in-person, and as it is becoming safe to return to workplaces there is a range of conflicting reactions from employees. This issue is set in the wider context of workers' rights and labor reform; the current labor shortage has led to a "Great Resignation" that has given workers a newfound power to push back against undesirable job conditions. Remote work offers employees a solution to many of these conditions, such as providing scheduling flexibility, better work-life balance, and independence. Additionally, the pandemic reframed what type of work is "essential", bringing into sharp contrast the need and undervaluation of retail and customer service workers. While these jobs are often considered low-skilled and underpaid, they are also both a necessity and not easily made remote, making them some of the most dangerous jobs to hold during a global pandemic.

Employers also have mixed feelings about the status of remote work. The pandemic is not over for many even as late as May of 2022, and making new postings remote may appeal to employers as remote work as a preemptive strategy against the lingering effects of new variants and outbreaks. For some, remote work affords an expanding pool of talent to hire from. For others, concerns about productivity, teamwork, and conventional conceptions of the workplace will incentivize a return to traditional in-person workplaces.

This context is especially relevant as graduating seniors. As we prepare to enter the job market, the available options are very different from what they were before the pandemic, especially in tech fields. The trends in remote work, which reflect the current change in mindset around workers' rights, also provide insight into the state of office and workplace culture, and even what it means to have a job. It is important to determine what changes will fade with the pandemic, and which are more permanent. As a result, the goal of this project is to use job data collected since the start of COVID-19 to predict the future of the job market, both in terms of how many remote vs. in-person jobs are available as well as the characteristics common to remote jobs versus non-remote jobs. The term 'remotable' has been used to describe how easily a job can be made remote, and our research attempts to quantify this quality based on similarities between existing remote jobs.

Our research does not use surveys or government data but instead uses data pulled off one of the most popular online job posting sites, and focuses on predicting future trends rather than explaining past data. Much existing research was concerned only with the immediate impact during the pandemic, but our research is motivated by post-pandemic inquiries and broader, lasting changes in the labor market.

We began this research with an expectation that the proportion would be higher for positions in the technology industry and other high-skilled sectors, and that the pandemic had accelerated the transition of jobs already trending towards remote or hybrid arrangements. However, it was unclear whether the overall proportion of the job market that was remote would increase or decrease, and to what extent.

2 Related Work

A number of papers have tracked employment for remote work during the pandemic. A study using nationally representative surveys from April to May 2020 found that half of people employed pre-COVID were working remotely, and higher rates of COVID were a predictor of more remote work (1). Other research tried to determine if remote work is in high demand based on evidence from job postings during COVID-19. Although all job postings decreased during the pandemic, the expectation is that job postings for non-remote work will have decreased less. However, while employment for "remotable" occupations remained steady through the pandemic, job postings in these occupations dropped (5).

Other research differed in its specific goals but used similar methods to investigate the job market. A study done in China collected job postings that involved blockchain technology, using keywords to identify relevant online job listings, and those listings to gain insight into employer demands when hiring (2). Another study done in the U.S. also created a dataset of online job postings and used keywords to study growth in the health information technology industry (8). Papers like these are important not only for insights into their individual topics, but for the methods they develop to process and analyze data from online job postings. Online job platforms are now increasingly the way employers and job seekers find each other and therefore can provide insight into the overall job market. In fact, those who search for work online are reemployed about 25% faster than comparable job seekers who do not, evidence for which refutes research from the 2000s purporting the ineffectiveness of the internet (4). This significant change requires new research methods to be developed in order to understand the evolving cycle of employment. Job sites are a significant part of this online search.

Relevant research not only studied the employment effects of the pandemic and the shift of job seekers to online resources, but the evolution of the workplace in parallel to broader societal changes. A paper from 2020 explores the phenomenon of "digital nomads", who are not tied to a physical workplace but instead work while traveling (3). This work arrangement and those similar to it is driven by the advent and rapid ubiquity of technology in modern life, the popularity and normalization of frequent travel and relocation for the lifestyle or more opportunities, and greater flexibility and precariousness of employment. Although travel was not a feature of the pandemic, the other circumstances of this work arrangement no doubt served as one of the few precedents and templates for the huge shift to remote work during COVID-19. Conceptualizing instances of remote work before the pandemic will help reveal and define the future of remote work after it.

If "digital nomads" epitomize the popular conception of remote work right before the pandemic, a paper from 1983 provides an overview of its emergence decades earlier. Margrethe H. Olson (6) examines the behavioral, organizational, and social issues regarding remote work and defines characteristics associated with it, many of which are echoed in modern research reviewed for this paper. We begin to see the adoption of technology to this end, and the communication networks necessary to this work. The details of technology and communication evolve, but many social characteristics ring true today, like minimum physical requirements, control over individual work pace, relatively low need for communication, high self-discipline, and a sought after skillset that afforded these workers the power to bargain for this flexibility (6). This difference for skillset falls in line with Brynjolfsson et. al. finding that "information" work and professional occupations were more likely to shift to working from home during the pandemic (1).

Surveying the reasons, circumstances, correlates and characteristics for remote work provides a baseline for evaluating the results of our research and inferring the future of remote work. Technology and the pandemic have been by far the largest drivers in the shift to remote work.

3 Dataset

The dataset used was collected by PromptCloud and DataStock, a web scraping company and a dataset vendor, respectively. The companies have collected 7 datasets of job entries from Indeed.com, each from a different period of time between May 2019 and September 2021. Full versions of the dataset containing multiple million listings each are sold through Datastock, and samples of each are available for free on Kaggle.com, which is what was used in this project. While each sample is listed as having 30,000 entries, two only had 30, bringing the total number of job listings available on Kaggle to 149,904.

Combined the datasets contain 58 variables, with some of the basic background variables including the job title, company name, listing text, and URL. Another subset of variables has multiple columns with different versions of the information; like all data filled in by humans, information in the job listings is sometimes left blank or incorrectly filled. Some entries have the city, state, and country filled out, while in other listings the state is listed as the city or left blank entirely. As a result, aside from the three columns listing the geographic location as collected, there are also inferred city, state, and country variables where data analysts have tried to fill in missing information. Another category represented by multiple variables is salary, with columns representing inferred salary low and high ranges, the inferred frequency of pay, and the inferred currency. One important variable for this analysis is a binary variable indicating if a job is remote. After preliminary preprocessing, it turns out the three datasets with entries before September 2020 do not have this remote variable, likely because Indeed.com didn't have a specific input for remote status. As a result those listings were removed from the dataset, so the final dataset of entries with a remote status variable contains 119,844 listings.

Additionally, for various analyses different formats or subsets were taken of the entire dataset. For example, only 30,234 listings had the salary included, so that subset of data was used for analyses regarding salary.

is_remote	post_date	category	inferred_salary_time_unit	mean_salary_range
false	2020-11-30	Administrative	hourly	17.0
false	2020-10-08	Healthcare	hourly	20.0
false	2020-10-30	Education Training	hourly	13.5

Figure 1: *The head of the salary dataframe.*

4 Methods

Our analysis involves two models; a linear regression to predict the proportion of jobs posted each day are remote, and a logistic regression to determine the remotability of a job. It was important to use the proportion of jobs and not the overall number to investigate to shift to remote work independent of the overall growth of the job market post-pandemic. To aid in our model, we used the pandas, NumPy, and scikit-learn python libraries. We used pandas for processing and analyzing our data, NumPy for arrays and matrices, and the scikit-learn Linear and Logistic Regression functions. All the data in our models was split into a test and a training set; 75% for the former, 25% for the latter.

Our linear regression model took in one feature, the date that the job was posted on, and output the proportion of remote jobs on that date. Initially, we implemented a simple regression that was linear, to flatly assess whether there was an overall upward or downward trend in job postings. Then, since we were only interested in one feature for this regression and the best model was clearly polynomial, we then implemented feature expansion to create a polynomial regression. We made the degree of this feature expansion a variable to serve as a hyperparameter in our search for the best model.

For training the model, we used our own implementation of gradient descent for the simple regression, and the scikit-learn implementation for the polynomial. The scikit-learn implementation was more efficient, which made it better for a more complicated model, but our own implementation was more configurable and lent easy access to the hyperparameters of iteration and alpha to control the step size.

To create a remotability factor, we wanted the model to take in features of a job posting and output a probability that the job is remote. Whether or not the job is actually remote, this can serve as a measure of how similar that job posting is to other job postings that are remote, and therefore an indication of how easily that job can be made remote, or how "remotable" it is. To create a simple model for this, we took three features of a section of the data set; the date posted, the category, and the salary. The category referred to the industry or sector of the job market, for example 'Sales' or 'Legal.'

5 Results and Evaluation

5.1 Using Linear Regression to Predict Proportion of Remote Jobs

The first model fitting process was a simple linear regression predicting the proportion of jobs from one day that were remote using the date. The best model will be determined by adjusting the hyperparameters of the gradient descent algorithm, and the model will be considered successful based on the behavior of the cost function over the iterations.

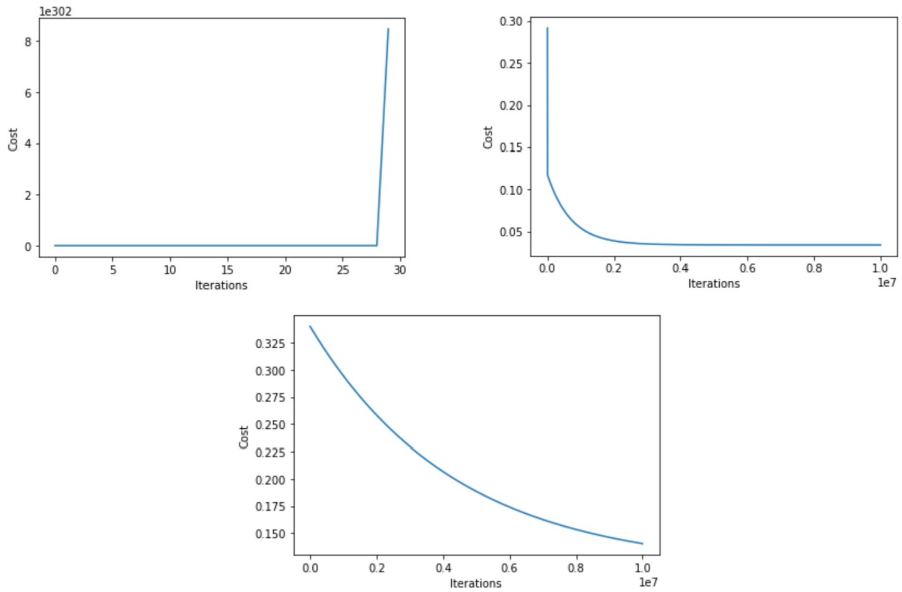

Figure 2: *Cost over 10,000,000 iterations for* α = $1\mathrm{e}-2$, $1\mathrm{e}-8$, *and* $1\mathrm{e}-14$ *respectively.*

Looking at plots of the cost over 10,000,000 iterations, as seen in figure 2, the ideal alpha value is around $1\mathrm{e}-8$. When alpha is equal to $1\mathrm{e}-2$, the steps taken to adjust theta are too large, overshooting the minimum of the cost function and diverging. At α = $1\mathrm{e}-14$, although the cost is moving towards zero it is taking steps that are too small and as a result the function will take too long to converge. At α = $1\mathrm{e}-8$, the cost function rapidly decreases and converges, making it the ideal choice.

The results of the simple linear regression indicate little variation in the proportion of jobs that are remote over time. Interpreting the slope, with the progression of every day, we predict that the proportion of jobs that are remote decreases by $5.98\mathrm{e}-6$. While this line does appear to be the best simple linear model to represent the relationship between date and proportion of remote jobs, it does not capture the variation seen in Figure 3. As a result, our next step is to use feature expansion to determine the best polynomial model to capture the trends in our data.

To fit a polynomial to our data, we used feature expansion. The first step is to determine the best degree of polynomial. A higher degree polynomial may fit the data better, but it risks being overfit to the data it was trained on. In other words, it is reacting to the "noise" of this specific sample as opposed to the overall trend between the variables. To avoid overfitting, we evaluate the performance

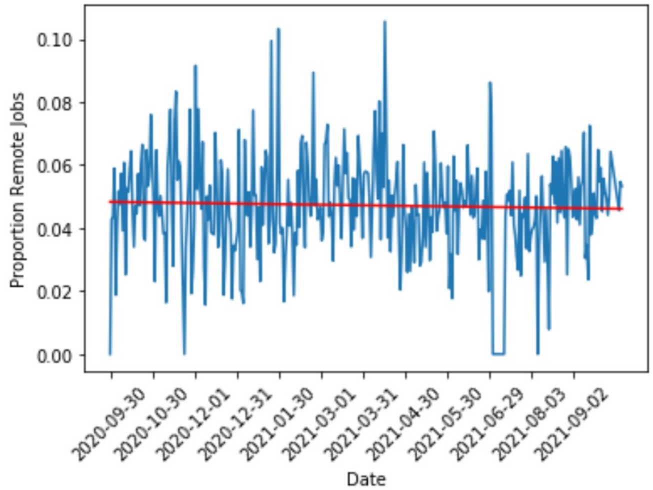

Figure 3: *A simple linear regression predicting proportion of jobs that are remote by day.*

of the model on a training and testing set. The ideal polynomial degree is the one with the smallest gap in performance between the training and testing dataset.

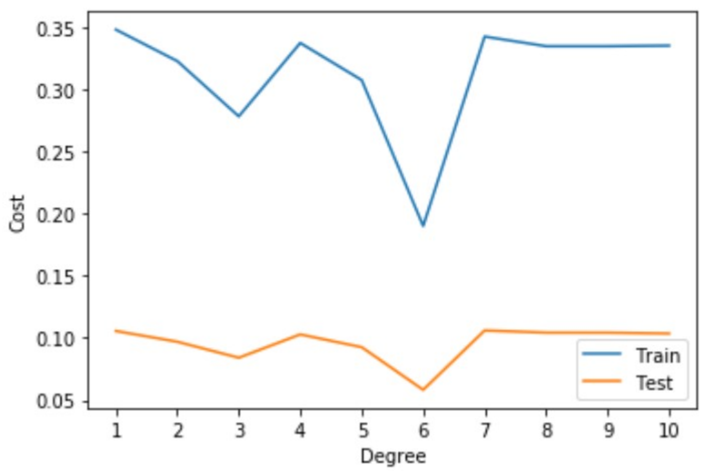

Figure 4: *Cost of the model by degree of polynomial for training and testing datasets.*

As seen in Figure 4, not only does a 6-degree polynomial have the lowest cost, it also has the smallest gap in training and testing data, indicating that it is the best choice. A 6-degree polynomial fit to the data also visually captures the variation of the proportion of remote jobs over date well, as seen in Figure 5.

The improved fit of the polynomial model reflects that the trend in remote job listings has not remained constant; our model indicates that there were the most remote jobs available in October 2020, March 2021, and August 2021, with dips in between.

6 Conclusion

Our simple linear regression model found that there was a slight decrease in the proportion of job listings that are remote between September 2020 and September 2021. In addition, a polynomial regression identified peaks in remote listings in October 2020, March 2021, and August 2021. Applying our model to predict dates outside the range of the available data, we predict that today the percentage of jobs listings that are remote will be 4.5%.

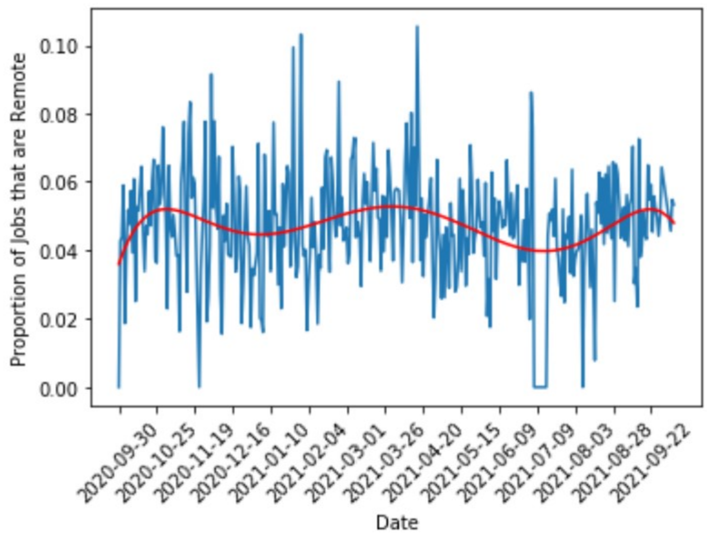

Figure 5: *A 6-degree polynomial fit to the daily proportion of jobs that are remote by date.*

Although we gained insight into the remote job market during 2020 and 2021, the limited data available to us means our predictions for future years are necessarily incomplete. Further research would include collecting data from the rest of 2021 and 2022 up until the present day. Portions of new data could serve as a training set to increase the predictive power of our model, and as a testing set to further evaluate the accuracy of our current model.

Future investigation of the nature of remote work should add additional features to our logistic model, including analysis of the textual job descriptions themselves. Additionally, because our model uses a multivariate logistic regression, additional classes can be added to the current remote binary to represent alternative work arrangements such as “hybrid” or “international”. This model can be used to guess at the remote status of a job when information is missing, to study the characteristics of past remote job listings, and to evaluate and output a “remotability score” indicating how easily a non-remote job could be made remote based on its similarities to already remote jobs.

Overall, predicting the future of remote work seems an attainable target, and we are confident that after the COVID-19 pandemic we now have enough data on remote job postings to inform future decisions about whether to make new jobs remote or not.

References

[1] Brynjolfsson, E., Horton, J.J., Ozimek, A., Rock, D., Sharma, G. & TuYe, H.Y. (2020) *COVID-19 and remote work: an early look at US data.* (No. w27344). National Bureau of Economic Research.

[2] Ge, C., Shi, H., Jiang, J. & Xu, X. (2021) Investigating the demand for blockchain talents in the recruitment market: evidence from topic modeling analysis on job postings. *Information & Management*, pp. 103513.

[3] Hannonen, O. (2020) In search of a digital nomad: defining the phenomenon. *Information Technology & Tourism* **22**(3):335-353.

[4] Kuhn, P. & Mansour, H. (2014) Is internet job search still ineffective?. *The Economic Journal* **124** (581):1213-1233.

[5] Morales-Arilla, J. & Daboin, C. (2021) Is remote work in high demand? Evidence from job postings during COVID-19. *ACM SIGCAS Conference on Computing and Sustainable Societies*, pp. 27–37. New York, NY: Association for Computing Machinery.

[6] Olsen, M.H. (1983) Remote office work: changing work patterns in space and time. *Communications of the ACM* **26**(3):182-187.

[7] Ozimek, A. (2020) The future of remote work. *Upwork* (https://www.upwork.com/press/releases/the-future-of-remote-work).

[8] Schwartz, A., Magoulas, R., & Buntin, M. (2013) Tracking labor demand with online job postings: the case of health IT workers and the HITECH Act. *Industrial Relations: A Journal of Economy and Society* **52**(4):941-968.