Understanding the Great Resignation

An Analysis of Job Ads and Attrition Data

Eli Jaffe, Frank Mo, Catalina Jaramillo, Paul Squires & Julian Togelius

NEW YORK UNIVERSITY
07/22/2022
Contents

1. Introduction
2. Datasets
3. Methodology
4. Conclusion
5. References
1. Introduction

The “Great Resignation”, a term credited to Anthony Klotz in his interview with Arianne Cohen (2021) for Bloomberg Businessweek, is reshaping the labor market in unexpected ways and with unprecedented speed. There is further disruption from factors such as the persistence of COVID-19, the war in Ukraine, and generational shifts in employees’ expectations. Over 12% of Gen Z employees left their job in the last 6 months while less than 5% of Millennials, less than 3% of Gen X, and less than 2% Baby Boomers did so, according to research by Revelio Labs (2022).

Due to these circumstances, we can say that today’s skill market is a “seller’s market”. There is strong organizational competition for skilled workers, which gives the candidate, or “seller” of skills, an advantage. In such a market, organizations must be strategic in their approach to human capital to ensure the availability of the right skills in the right positions. The ability to forecast labor market behavior is key to making better decisions.

Forecasting such a dynamic market is no easy task. However, the development and proliferation of forecasting tools allow us to make ever more complex models. The NYU Human Capital Analytics Lab, in collaboration with Henning Seip at Candogram, set out to apply these emergent methods in the context of the Great Resignation.

2. Datasets

In order to learn better forecasting models, the researchers used two data sources: Candogram (2022) and Job Openings and Labor Turnover Survey (JOLTS) from the U.S. Bureau of Labor Statistics (BLS) (2022).
Candogram

Candogram provides job market education to students in high school, college, and bootcamps. It turns millions of live job postings into an interactive learning experience where students track job demand and required skills as they invest in their education. These job postings come from a daily download from corporate websites of about 10,000 US employers conducted by a public-private partnership. Candogram extracts key data points from the ads, including quantity, related skills, job family, employer, industry, and location. These fields are used in further analyses and focus on technical jobs and skills.

JOLTS

JOLTS is a traditional source of labor market data: it is frequently used by companies to understand industry retention, identify training & development needs, and perform labor force planning. It is also a source for research applications. JOLTS is published by the BLS and provides data on total employment, job openings, hires, quits,
layoffs & discharges, and other separations. Data is obtained via a monthly survey of more than 20,000 business and government entities and is released on a monthly basis.

3. Methodology

In order to combine data from both sources, we trained unique AI models using data from companies within two groups of industries. The groupings were determined based on company classification according to the BLS North American Industry Classification System (NAICS) industry designation:

- **Knowledge Industry Grouping**: includes employees of organizations in the Information, Professional, Scientific, and Technical Services, Finance and Insurance, and Management of Companies and Enterprises industries (in other words, companies within NAICS codes 51000, 520099, and 540000).

- **Service Industry Grouping**: includes employees of organizations in the Health Care and Social Assistance, Accommodation and Food Services, and Educational Services industries (i.e., NAICS codes 600000 and 720000).

The primary objective of our research was to develop forecasting models using advanced AI methods that allow us to understand the behavior of the job market by predicting the expected number of jobs to be posted in the future. As a secondary goal, we were interested in how these forecast models could be applied to understand the labor market demand for technical skills related to the predicted job postings.
A forecast is a prediction of the future trend of a variable, based on past and/or present data. This can be done using the past behavior of the same variable (time series) or past/present levels of external variables that have an impact on the variable of interest (regression approach). A time series consists of data points that are indexed in a timeline: any variable that is presented as a function of the elapsed time is considered a time series (Hayes, 2021).

The team considered different forecasting methodologies to build the models:

- **Holts-Winter**: Identifies the average, the trend, and seasonal changes in the time series behavior of the data. By combining the effects of these three elements, this model predicts the next value for the variable in the time series (SolarWinds, 2022).

- **SARIMA**: Seasonal Auto Regressive Integrated Moving Average. This uses parameters to calculate the prediction based on autoregression, differentiation, and moving averages. It also learns the cycle of seasonality. (Brownlee, 2019).

- **Prophet**: Bayesian-based curve fitting, developed by Facebook (now Meta). Prophet combines the effects of yearly, weekly, and daily changes and adds holidays’ impact on a time series to fit the model (Taylor & Letham, 2018).

- **Silverkite**: Based on conditional mean (time, day, week, etc.) and volatility/error; by LinkedIn. This combines a conditional mean model with a volatility model that learns variations from specific characteristics of the time series such as day of the week (Lewinson, 2022).

These four methods use time series as input and focus on identifying patterns including trends (growth, stability, etc.) and seasonal changes (monthly, weekly, or any kind of cycle observed in the data) to predict the future output. Some models, like Prophet and Silverkite, allow for additional variables to be fed to the model in order to improve the prediction accuracy.

**Technical Methods**

The Prophet and Silverkite models were applied to the Candogram job posting dataset using a three month forecast window. The Mean Absolute Error (MAE) was chosen as the accuracy measure to assess model performance. Job posting data
from April 2019 to March 2021 was aggregated monthly and split into a training set of 21 months (April 2019 until December 2020) and a test set of 3 months (January 2021 through March 2021). The Prophet and Silverkite models were then trained using (1) the standalone Candogram job posting time series and (2) the Candogram data with the external JOLTS data. To create the best forecast, eight different sequential lags were calculated on the JOLTS data ranging from one month until eight months prior to the target prediction date. Each of the 8 sequential lagged time series was paired with the Candogram data and used to train the Prophet and Silverkite models. The MAE between the predicted job postings from the model and the count of actual job postings observed was then calculated for each model. The model with the lowest MAE testing error was selected for each industry grouping. A diagram of these steps is presented below.

**Diagram:** The analytical process to determine the best forecasting model for each industry grouping (Knowledge Industry and Service Industry). Knowledge Industry and Service Industry data was collected from Candogram and combined with JOLTS data at sequential lags. The Prophet and Silverkite models were then trained on each combination of Candogram + JOLTS data and the model of each industry grouping that had the smallest Mean Absolute Error (MAE) was selected. For both industry groupings, models that included JOLTS data performed better than models that did not. However, the best lag to use differed between industry groupings.
4. Results

Industry Trends

The Prophet model that included the JOLTS lagged Quits data performed better than the models with the Candogram job posting data alone. In other words, we gain information about the number of future job openings by looking at the historical quit rate. Furthermore, the most useful sequential lag differed between industry groupings. For the Knowledge Industry Grouping, quit data from 6 months ago (lag 6) was most helpful for predicting the number of job openings whereas for Service Industry Grouping, quit data from only one month prior (lag 1) gave the best estimate.

There is an important constraint when working with BLS data. The BLS does not publish JOLTS data every month, so although the Candogram with JOLTS lag 1 model performed best for the Service Industry Grouping, there are real world limitations on its use. The pattern we found for the Service Industry Grouping suggested that as the size of the lag approached 1, the accuracy of the job posting prediction got better. The Prophet model with a 3 month lag had similar accuracy to that of the model that did not include JOLTS data at all. Balancing accuracy with utility of forecast, we selected the Service Industry Grouping model that incorporated JOLTS data at a 3-month lag. Figure 1 shows the relative error for each model.

The difference in meaningful lags between quits and job postings suggests that there are industry group-specific differences in labor market dynamics. The Knowledge Worker job vacancies are filled more slowly when talent leaves while the Service Worker job vacancies appear to be filled more rapidly. Therefore, it is important to consider the industry grouping of interest when incorporating quits information into job posting forecasts.
model accuracy improved beyond the model trained on the standalone Candogram data. However, the ability to forecast into the future became more limited, so JOLTS lag 3 was chosen.

As stated above, the performance of the models was defined as the error between the predicted job postings and the actual job postings for all observations. To see how each of the models performed, we looked at the historical job count over the duration of our data series along with the predictions of each model (Figure 2). We can see that, over time, the top performing models align with the actual values much better than the less well-fitting models.

The models provide a sense of where the market will go for each industry grouping. The model predicts that the demand in the Knowledge Industry Grouping will continue to increase in the upcoming months and then may level off slightly. For the Service Industry Grouping, our forecast indicates that job demand will remain at a similar level to current demand.
Figure 2. Actual number of job advertisements (black line) compared to the Prophet and Silverkite models with and without Quits data for the Knowledge Industry Grouping (top) and the Service Industry Grouping (bottom). The better performing models more closely align with the actual counts of job postings and add credibility to the forecasts, that is, the job posting predictions after the dotted line.
The relationship between quits and job postings shows promise in that quits data can be seen as a leading indicator of job postings. HR leaders may be able to get an early sense of the direction in which the labor market is going by looking at the trend in quits for their specific industry. Similarly, job seekers may be able to leverage this information to plan their job search or decide to renegotiate with their current employer.

**Skills Trends**

To gain a greater understanding of labor market trends, we analyzed the skill sets demanded by employers in the Knowledge Industry Grouping and the Service Industry Grouping. In the Candogram dataset, there are six categories of technical skills: Programming Languages, Database tools, Back End Frameworks, Front End Frameworks, Operating Systems, and ERP Systems. Figure 3 shows the time series of technical skill sets listed in job postings for the Knowledge Industry Grouping. Database skills are in the greatest demand. However, there is an upward pattern of demand in all six skill categories. This indicates that for jobs in the Knowledge Industry Grouping, the demand for these technical skills may be related, notwithstanding the fact that the demand for database skills is rising faster than that of other skills, and the gap between them is widening.

**Figure 3**

![Figure 3. Daily demand for technical skills in the Knowledge Industry Grouping from March, 2021 until March, 2022.](image-url)
In contrast, the demand for different skills within the Service Industry Grouping is more variable (Figure 4). Even though the trend for skills demanded by companies in the Service Industry Grouping is also increasing, the trends for the separate technical skills have some different features than that of the Knowledge Industry Grouping.

**Figure 4**

![Figure 4. Daily demand for technical skills in the Service industry Grouping from March, 2021 until March, 2022.](image)

As can be seen in Figure 4, the demand for certain skills, such as for front-end frameworks and operating systems, has shown signs of decline. Additionally, the separate demand trends for the six sets of skills are not parallel, that is, they cross (e.g., Front-End and Operating Systems). This did not occur in the Knowledge Industry Grouping. As with the Knowledge Industry Grouping, the demand for database knowledge still outpaces every other skill.

We applied the Prophet modeling technique to the skillset time series for each industry grouping to forecast where demand might go. We chose the Prophet method because, overall, Prophet and Silverkite performed better than the traditional time series model. The forecasts for each industry grouping are shown in Figures 5 and 6.
**Figure 5.** Prophet model forecast for the demand for technical skills in the Knowledge Industry Grouping from March, 2021 until March, 2022.

**Figure 6.** Prophet model forecast for the demand for technical skills in the Service industry Grouping from March, 2021 until March, 2022.
The skill demand forecast in the Knowledge Industry Grouping follows the job posting trend formerly described. There is steady growth in future demand for all six technical skill sets. However, the skill demand forecast for the Service Industry Grouping shows something different. The Prophet forecasting model predicts that only the database skill set will experience steady growth while the demand for other skill sets will decrease or remain unchanged. It can be inferred that job postings within the Service Industry Grouping may not require the same breadth of technical skills in the future.

Data from the JOLTS survey were not included in the skills demand forecasting model due to limitations in the length of historical information collected. The available skills data began in March, 2021. Aggregated to a monthly frequency, the time series was too short to make precise forecasts. As skill data continues to be collected, the JOLTS data stand to be a promising addition to the model.

4. Conclusion

There are three important implications of the present research. First, models that forecast job postings are improved when resignation data is added. Therefore, when considering labor market demand, corporate leaders and individual employees will make better decisions by using forecasts that include publicly available data, that is, JOLTS quit rates.

Second, there appear to be meaningful differences in how companies react to resignations based on industry membership. Forecasters may be better served to look at industry-specific quit rates than macro-level trends. The current study investigated Knowledge Industry and Service Industry groupings; future research may extend our understanding of these dynamics in additional industries.

Third, while the demand for technical skill sets across Knowledge and Service Industry groupings have some similar patterns, important differences exist. This research predicts that greater change is ahead for the Service Industry than Knowledge Industry groupings.
The current study uniquely combined independent sources of human capital data (Candogram data and BLS data). The complementarity of these data enhances the validity of the findings. Other independent sources of big data may be leveraged in the future to further increase prediction accuracy. Additionally, more powerful models (such as Prophet and Silverkite) show promise as tools for researchers and analysts alike.

For applied practitioners, providing HR leaders with accurate forecasts of the labor market demand for critical skills enables them to act preemptively and make better talent management decisions which, among other benefits, may reduce resignations. For employees, knowledge of the demand in the labor market for specific skills can enable them to make better career planning and development decisions and may reduce their likelihood of resignation.
5. References

—


