MARCH 2022

Unpacking The Great Resignation

~Examining Key Drivers of Attrition Across U.S. Industries~

Reyada Atanasio   Maryama Diaw   Alexis Fennoy

O. Justina Molokwu   Tabitha Mundia

DS 4A

COHORT 3
TEAM 12

TABLE OF CONTENTS

PROJECT OBJECTIVES

Our team wanted to identify what industries have been experiencing the highest attrition rates.

We wanted to compare emerging trends in quantitative data from labor force statistics with sentiment analysis derived from web scraping social media to get to the root of employee attrition so employers can build out more effective retention strategies.

Given that the ultimate goal of many organizations is retaining the talented employees they recruit, our project focused on analyzing how retention and attrition rates by industry are related to prevalent factors (i.e., salaries and wages, health care benefits, etc.) that affect work in those industries.

Key Focus Areas

Retention & Attrition Rates by Industry

Sentiment Analysis by Web Scraping

PROJECT STATEMENT & CONTEXT

COVID-19 has driven unprecedented shifts in what work looks like, how work gets done, and the working standards that employees expect from their employers. The rate of employee resignations in the United States broke records in August 2021 with 4.3 million Americans resigning from their positions (New York Times, 2021).

This record was broken yet again just one month later with another 4.4 million employees resigned in September 2021 (CBC News, 2021). This phenomenon, which is largely being referred to as “The Great Resignation”, is gaining momentum and has created a major dilemma for U.S. organizations.

More and more, the number of job openings are exceeding the number of workers looking to fill them, leading to vacancies that in turn put strains on organizational productivity.

PROJECT STATEMENT & CONTEXT

A 2018 survey of employees who chose to leave their organizations cited money and career growth as their primary incentives (SHRM, 2018).

However, a recent suvey of employees who chose to leave their organizations during the pandemic cited lack of flexibility, instances of discrimination, their contributions and ideas not being valued, insufficient benefits, and feeling their well-being was not being supported by their organization as major determining factors.

The contrasting motivations for attrition we are seeing in these pre- and post- pandemic surveys most likely signifies a change in the overall values of the workforce and therefore should also signify a change in the methods organizations utilize to retain current and prospective employees.

In a recent survey of skilled workers, 57% of respondents said they'd consider taking a new role in the coming year (Hollon, 2021). While the factors influencing these responses are not clear, the percentage of respondents willing to part ways with their organizations is alarming.

The results of this survey paired with the record breaking quit rates we are already seeing suggests attrition will continue to be a major dilemma for organizations moving forward.

The mass exodus of workers from the U.S. workforce has severe implications for:

- Domestic and global supply chains;
- Growth and profitability of businesses;
- Budgeting for monetary and non-monetary federal aid;
- The U.S. economy.

PROJECT STATEMENT & CONTEXT AUDIENCE

While many can benefit from the completion of this research, three key audiences that may find this research particularly valuable are:

01 — Government

The government is responsible for ensuring its citizens are able to attain gainful employment, without which they require federal support.

Findings from this analysis can aid federal & state government agencies in introducing structures, policies, and practices (e.g., raising the federal and/ or state minimum wage, implementing restrictions on work hours, toughening compliance laws) that help alleviate the concerns driving off workers.

02 — Employers & Recruiters

Employers can use this data to identify key factors of attrition and develop strategies that will enable them to retain employees and attract more diverse talent. In a similar vein, recruiting agencies can leverage this data to inform their recruiting protocols and help organizations attract and retain top talent.

03 — Economists

Economists–both professional and academic–study such phenomena and can utilize this information to inform predictions regarding future labor market trends. This data will provide additional perspective on and interrogation of trends related to subpopulations that are largely underrepresented in the literature.

The results of our data-driven research can benefit organizations personnel and profit goals. Through understanding the factors that drive the Great Resignation, employers are better able to develop strategies that offer staff conducive work arrangements without compromising corporations' strategic objectives.

DATA ANALYSIS
DATA WRANGLING & CLEANING

Dataset 1: Bureau of Labor Statistics (BLS)

We analyzed data from BLS, which contained workforce statistics on 10 industries across the United States and the District of Columbia.

We explored the data available from 2010 - Q3 2021 to determine which industry trends correlated with increased separations.

Construction	Education and Health Services	Financial Activities	Information	Leisure and Hospitality
Manufacturing	Professional and Business	Trade, Transportation and Utilities	Natural Resources and Mining	Other services (except public administration)

Variables

We identified ten variables of interest to analyze across industries:

- **Unemployment rate**: Number of unemployed individuals as a percentage of the total labor force
- **Job openings**: All positions that are available and for which an organization is actively hiring
- **Hires**: All part-time and full-time additions to the payroll
- **Separations**: All employees separated from the payroll as a result of quits, layoffs, discharges, or other separations (e.g., retirement)
- **Retirement benefits**: Percentage of employees with access to and participation in employer-provided retirement benefits
- **Health care**: Percentage of employees with access to and participation in employer-provided health care plans
- **Paid vacation:** Percentage of employees with access to paid time off
- **Paid sick leave**: Percentage of employees with access to paid sick leave for any quarantine or isolation orders
- **Total compensation**: Total compensation cost per hour worked for private industry workers

Data Cleaning

We utilized the following data cleaning processes prior to our analysis:

- Pulled data from BLS for each of the variables listed across each industry of interest
- Exported monthly averages of each variable from 2010- Q3 2021
- Standardized rows and columns across industries

DATA ANALYSIS
DATA WRANGLING & CLEANING
Dataset 2: Twitter

Twitter is a global platform with 396.5 million users, 206 million of which access the platform once per day (Backlinko, 2022). This provided us with a unique opportunity to gain significant insight into the conversations and sentiments of professionals regarding their relationship with their employer, employment status, resignation, hiring and the effects of the Great Resignation.

To determine if shared sentiments across job seekers explained some of the observed shifts in the labor market, we utilized Twitter's API to perform web scraping. This resulted in qualitative information on the Great Resignation through use of the following Twitter features- Tweets, Users, Spaces and Polls.

SCRAPED 44,668 TWEETS FOR THE FOLLOWING KEYWORDS:

Resignation - 25,262 tweets
Quit my job - 9,762 tweets
Great Resignation - 4,811 tweets
Toxic job - 1,861 tweets
People quitting - 1,574 tweets
Leaving jobs - 1,398 tweets

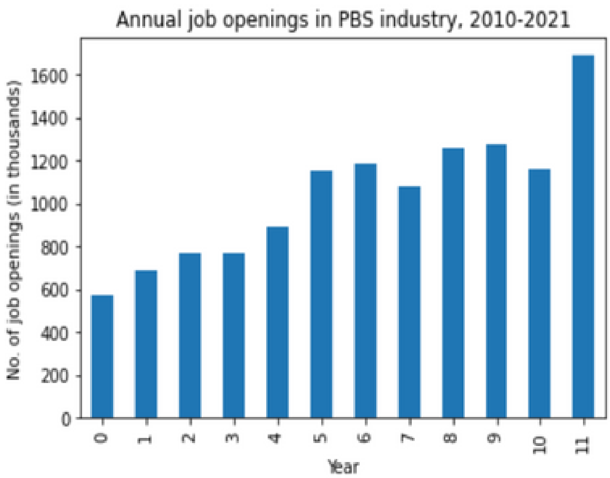
Data Cleaning

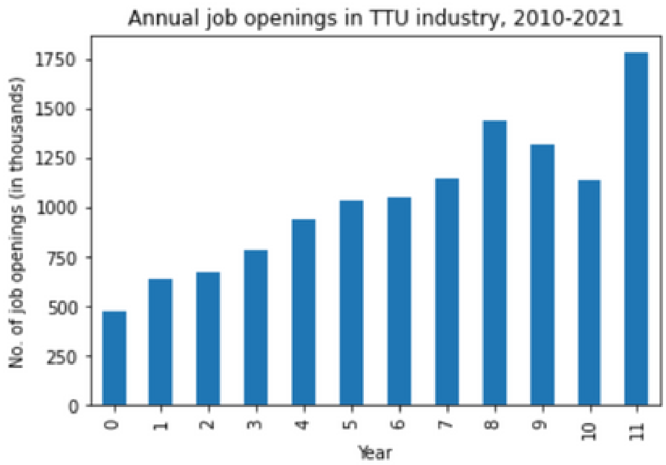
We utilized the following data cleaning processes prior to our analysis:

- Initial web scraping was carried out with Twitter using Pandas and Tweepy packages scraping the conversations with the key words “Great Resignation” and “quit my job”.
- Exported the results from “Great Resignation” and “quit my job” into Excel
- Cleaned and imported results to MonkeyLearn to view as word clouds

DATA ANALYSIS
EXPLORATORY DATA ANALYSIS

The first step in the EDA process was to understand general workforce trends across each industry. For example, figures 1 and 2 are simple bar plots to understand the distribution of job openings between 2010 (symbolized by 0) and 2021 (symbolized by 11). Both the Professional and Business Services (PBS) and Trade, Transportation, and Utilities (TTU) display a similar upward trend in job openings between 2010 and 2021; however, the volume of job openings is larger in the TTU industry. Furthermore, while job openings appear to have plateaued between 2018-2020 in the PBS industry, they dipped in the TTU industry before spiking in 2021.





Figures 1 and 2: Job openings in Professional and Business Services (left) and Trade, Transportation, and Utilities industries (right), in thousands, 2010-2011.

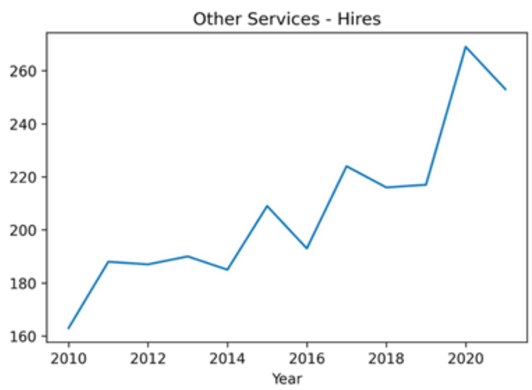


Figure 3: Hires for those working other services, which BLS defines as ranging from advocacy to personal care services appears to have been increasing since 2014. Average hires began to decline between 2020 and 2021 by 16,000.

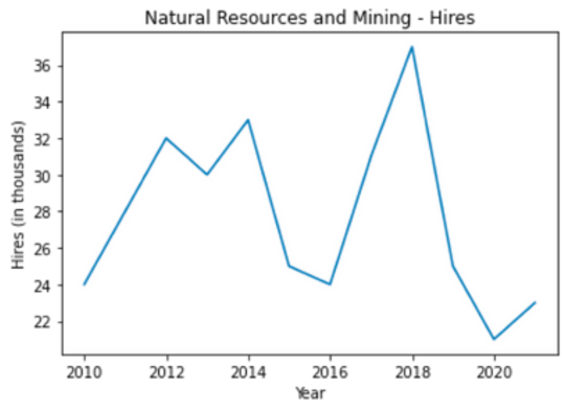


Figure 4: The Natural Resources and Mining industries experienced fluctuation between 2010 and 2010 with 2018 having the most hires (avg = 37,000).

DATA ANALYSIS
EXPLORATORY DATA ANALYSIS

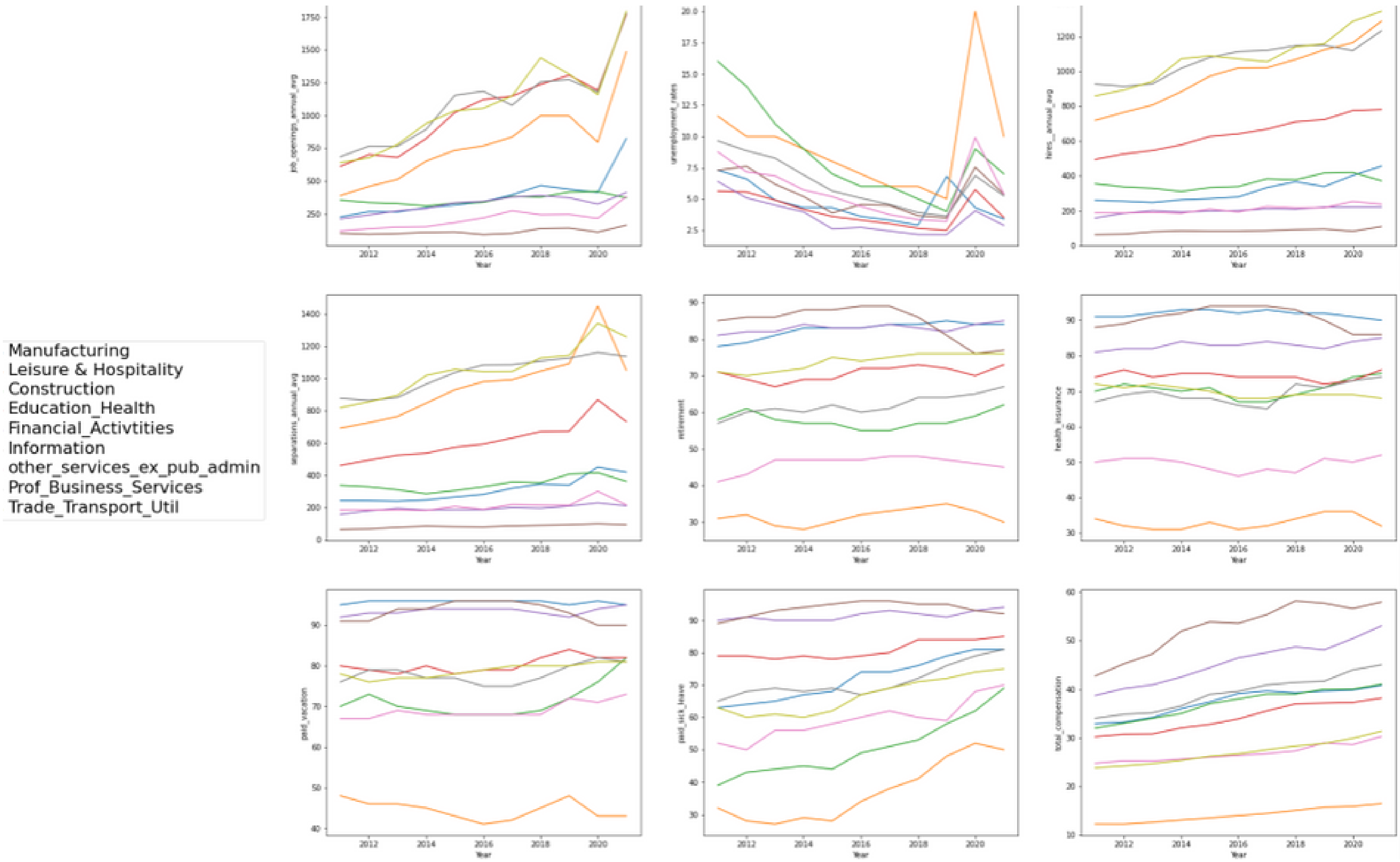


Figure 5: Trend analysis of the behavior of key aspects of employment over time across industries.

As we hypothesized, we observed that unemployment rates were the highest in 2020 due to the pandemic, and consequently, job hires and openings were high in 2021. Interestingly, job openings and hires in the trade, transport and utilities industry were high in 2021. This indicates that as the pandemic slowed down, people became more comfortable with travelling, as opposed to visiting restaurants and hotels, hence the observed faster hire rate.

Jobs in the education and health sectors were increasing over the years up until 2020. In the information industry, job benefits decreased during the pandemic. Further information is required to understand why.

DATA ANALYSIS
EXPLORATORY DATA ANALYSIS

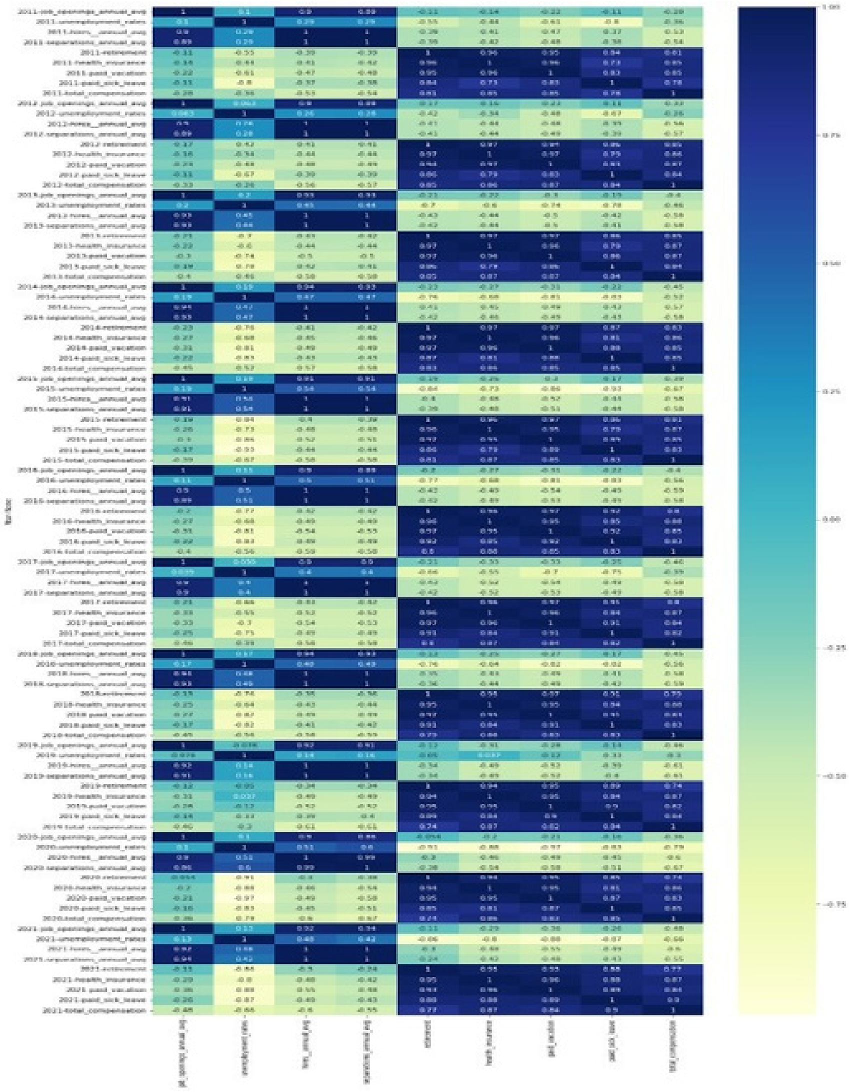


Figure 6: Correlation matrix analyzing variables of interest over time.

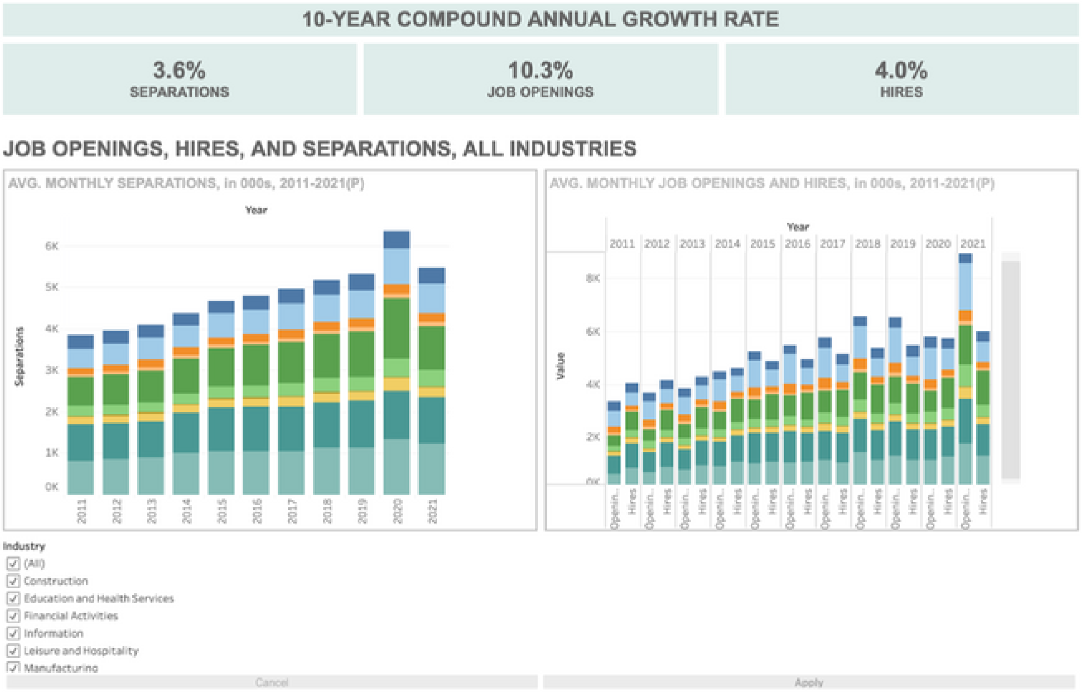
Unsurprisingly, we found strong correlations between job openings, hires and separations. Interestingly enough, there is a fairly weak correlation between unemployment rates and job openings and a fairly strong correlation between hires and separations. These correlations (unemployment vs hires and separations) were observed in 2018 and then they weakened in 2019, thereafter becoming the strongest in 2020 and 2021. Weak correlations were consistently observed between job openings, hires, separations and job benefits over the years. There are strong negative correlations between unemployment rates and the various benefits over the years and the strongest negative correlations were observed during the COVID-19 pandemic (2020-2021). Additional analysis is needed to better understand these findings.

DASHBOARD & RESULTS
USE CASE + DATA ENGINEERING

Figure 7: Dashboard displaying industries analyzed, context and snapshot of analysis, and summary statistics across key job turnover indicators

The resultant dashboard is a culmination of the data we found most insightful and explanatory of the workforce trends and key drivers that have ultimately led to the Great Resignation. After a brief overview of the dashboard, which provides context on the impetus for the data analysis and orients the user to how to interact with the dashboard, we display the ten-year (2011-2021) compound annual growth rates (CAGRs) for our three key labor turnover indicators: separations, job openings, and hires. As shown in Figure 7, job openings grew at an astonishing 10.3% CAGR during this time period. While some of this growth can be attributed to job creation, a closer look at the volume of average monthly job openings per year paired with qualitative information regarding the state of the workforce suggests a large share of openings in 2021, the highest of any year, was due to separations.

DASHBOARD & RESULTS - BLS USE CASE + DATA ENGINEERING



Figures 8 & 9: Dashboard charts displaying average monthly separations from 2011-2021 (left) and average monthly openings and hires during the same timeframe (right)

Further down in the dashboard, there are two interactive visuals: 1) a stacked bar chart displaying average monthly separations, in thousands, between 2011-2021 and 2) a clustered stacked bar chart illustrating job openings and hires, in thousands, between 2011-2021. As seen in Figure 8, the first chart exhibits a peak in separations in 2020, as the initial stage of the COVID-19 pandemic shifted traditional compositions of employment and employee perceptions of their relationship to work. Unsurprisingly, the industries that experienced the highest attrition rates–particularly in 2020–were the Leisure and Hospitality and Trade, Transportation, and Utilities industries; wage stagnation, health and safety concerns in the midst of the pandemic, and resulting job dissatisfaction are likely key motivations for attrition across these selected industries.

In the second chart, we see openings have historically outpaced hires–dating back to 2015–however, there is a significant disparity between openings and hires in 2021. Said differently, demand for employees is greater than available supply.

Viewers of the dashboard are able to isolate specific industries of interest using the filter below the charts. Furthermore, they are able to get a closer look at the data using the ribbon in the top right corner of the dashboard (see Figure 7). By making these charts interactive, our intention is to allow diverse stakeholders to tailor the data to their use cases and see the connections between different job turnover indicators in their target industries.

DASHBOARD & RESULTS - BLS USE CASE + DATA ENGINEERING

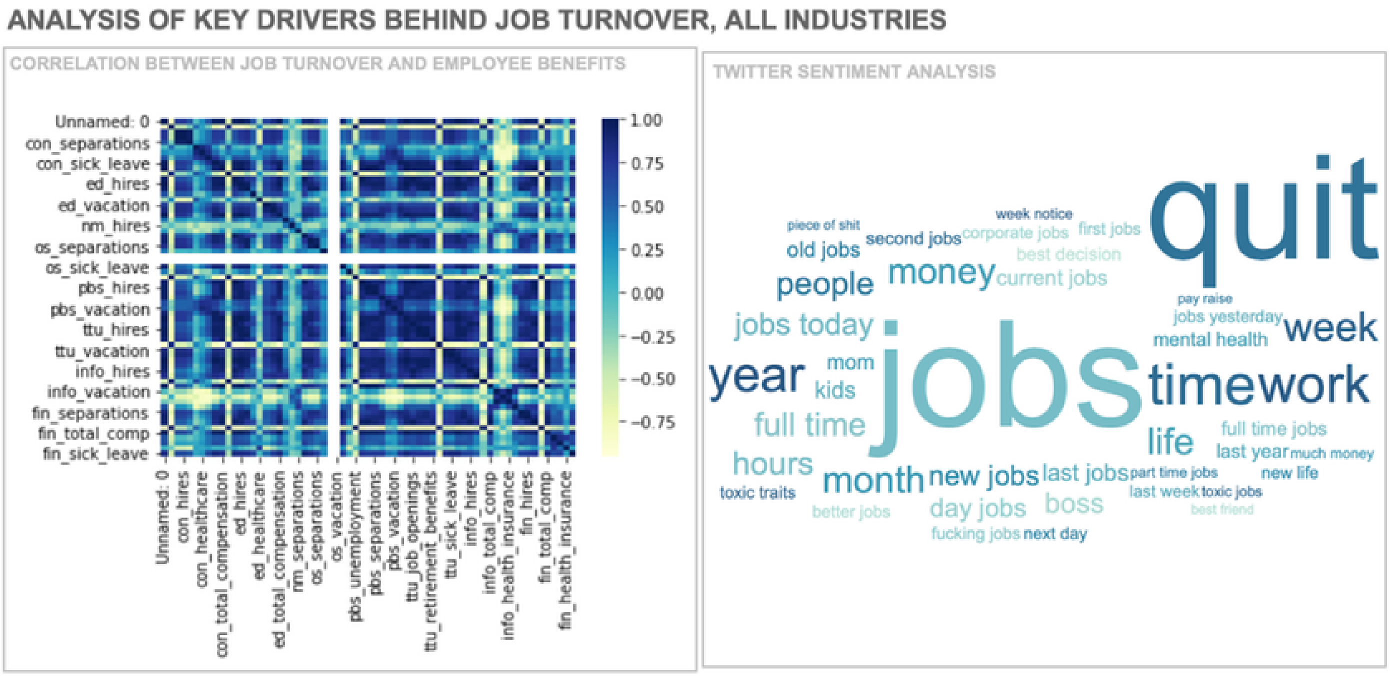


Figure 10: Results of correlation coefficient analysis (left) and Twitter sentiment analysis (right)

Lastly, we ran a correlation coefficient analysis to determine if there is a relationship between our three job turnover indicators and employee benefits (i.e., retirement benefits, paid vacation, sick leave, compensation) made available across industries. Our findings conclude there is a correlation between job turnover and benefits packages; however, the strength of the correlation varies across industries. For example, there is a strong correlation between these variables within the Professional and Business Services industry, but the correlation between benefits and job turnover in the Trade, Transportation, and Utilities industries is much weaker. See Figure 10 for further details of the analysis.

The Twitter sentiment analysis further corroborates findings from the quantitative analysis of BLS data. The frequency in citations of terms such as “jobs”, “quit”, “toxic jobs” when conducting a keyword search on “quits” highlights people’s dissatisfaction with their jobs, with the use of words such as “money” and “mom” signaling reasons why. The results of the Twitter sentiment analysis are expounded upon below.

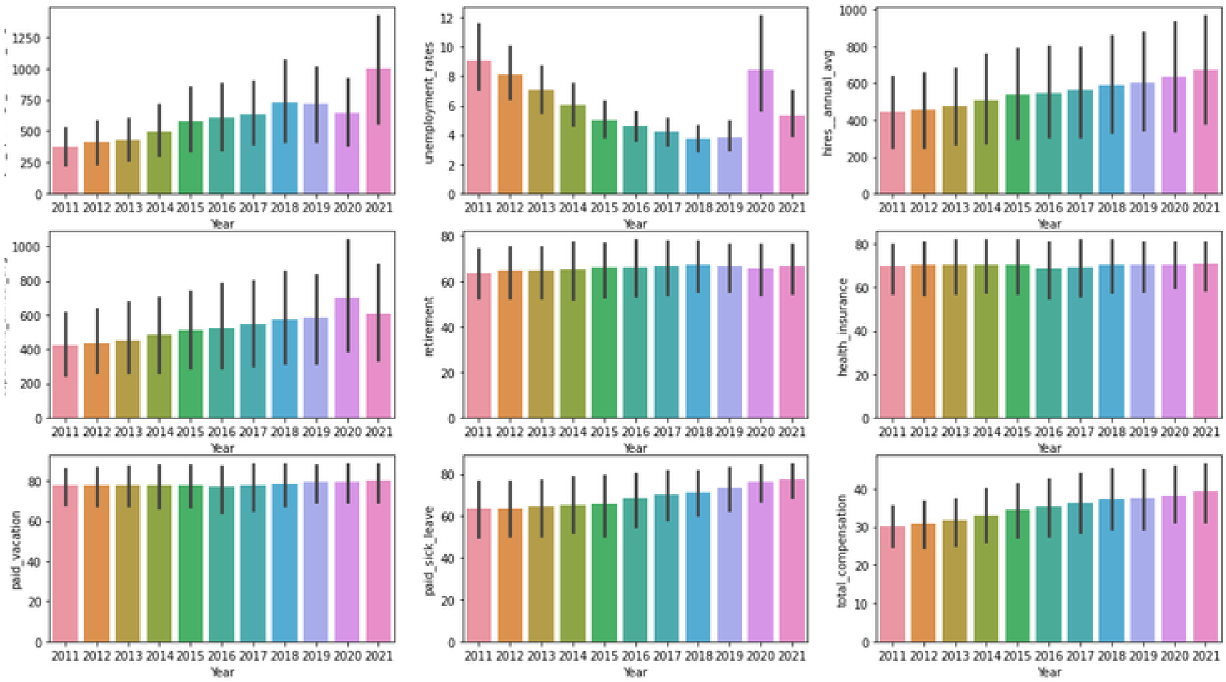
DASHBOARD & RESULTS - TWITTER USE CASE + DATA ENGINEERING

The initial web scraping process was carried out with Twitter using Pandas and Tweepy packages. Further analysis was done on the conversations with the key words “Great Resignation” and “quit my job”. The web scraping results from these phrases were exported to Excel where they were cleaned to extract only the sentiments, removing any date-time stamps and user identifiers. Then, this cleaned data was exported as a .txt file and uploaded to MonkeyLearn software to produce the word clouds for "quit my job" on the left and "Great Resignation" on the right below:

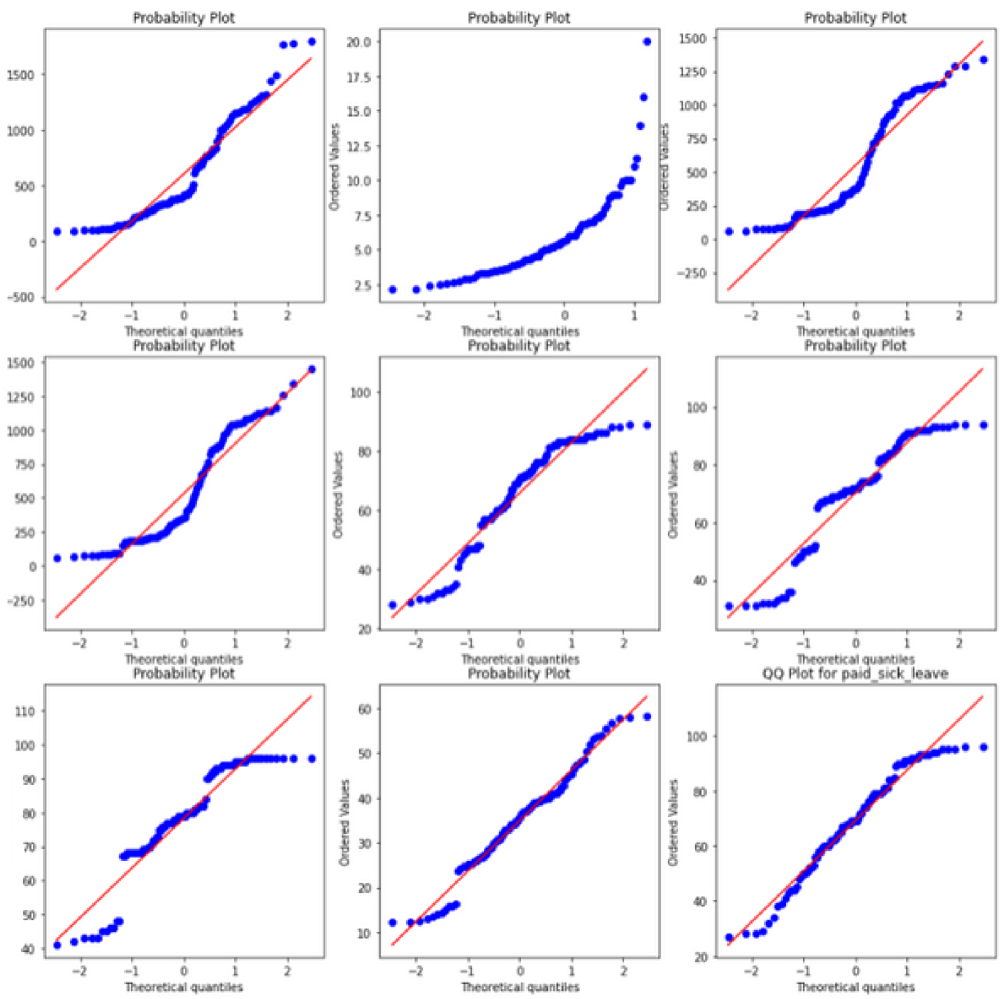
We found similarities to our quantitative BLS data analysis in references like **equal pay** and **benefit companion**. Additionally, we saw references that alluded to employee experiences in mentions of **toxic culture, toxic traits, kids, hours** and **mental health**. There was also specific mention of other industries such as **teachers** and **insurance** that did not have strong turnover statisticswhen examining the BLS data. This indicates that there may be more industry specific nuances worth further exploration to truly make impactful change for employees.

STATISTICAL ANALYSIS

We analyzed the distributions of our variables of interest from our BLS datasets:



Based on the bar plots, most variables are almost uniformly distributed apart from the unemployment rates indicating the key aspects of employment have remained fairly stable over the years apart from slight jumps observed in 2020 for separations, job openings, unemployment rates. To confirm, we created the below QQ plots:



We observed that apart from unemployment rates, the other variables are more or less normally distributed so we only transformed the variable unemployment rates.

STATISTICAL ANALYSIS

To better understand the key drivers of unemployment rates, we fit a simple linear regression model which allowed us to study these over the years, based on our variables of interest from the BLS. The results are below:

Dep. Variable:	transformed_unemployment_rates	R-squared:	0.713
Model:	OLS	Adj. R-squared:	0.653
Method:	Least Squares	F-statistic:	11.93
Date:	Sun, 27 Mar 2022	Prob (F-statistic):	4.79e-14
Time:	14:40:41	Log-Likelihood:	-1918.7
No. Observations:	88	AIC:	3869.
Df Residuals:	72	BIC:	3909.
Df Model:	15		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	4.904e+09	3.8e+09	1.291	0.201	-2.67e+09	1.25e+10
industry[T.Education_Health]	-9.041e+08	7.95e+08	-1.138	0.259	-2.49e+09	6.8e+08
industry[T.Financial_Activtities]	1.683e+09	1.27e+09	1.322	0.190	-8.55e+08	4.22e+09
industry[T.Information]	1.069e+09	1.5e+09	0.711	0.480	-1.93e+09	4.07e+09
industry[T.Leisure & Hospitality]	-5.697e+08	2.2e+09	-0.259	0.796	-4.95e+09	3.81e+09
industry[T.Manufacturing]	7.229e+08	1.15e+09	0.629	0.531	-1.57e+09	3.01e+09
industry[T.Prof_Business_Services]	1.65e+08	1.57e+09	0.105	0.917	-2.97e+09	3.3e+09
industry[T.Trade_Transport_Util]	-2.048e-05	2.32e-05	-0.882	0.381	-6.67e-05	2.58e-05
industry[T.other_services_ex_pub_admin]	6.756e+08	1.37e+09	0.492	0.624	-2.06e+09	3.41e+09
job_openings_annual_avg	1.519e+06	1.23e+06	1.233	0.222	-9.37e+05	3.98e+06
hires__annual_avg	-2.213e+07	4.67e+06	-4.742	0.000	-3.14e+07	-1.28e+07
separations_annual_avg	2.173e+07	2.31e+06	9.391	0.000	1.71e+07	2.63e+07
retirement	-7.526e+07	4.88e+07	-1.541	0.128	-1.73e+08	2.21e+07
health_insurance	1.744e+08	6.29e+07	2.773	0.007	4.9e+07	3e+08
paid_vacation	-1.688e+08	6.35e+07	-2.657	0.010	-2.95e+08	-4.22e+07
paid_sick_leave	5.14e+07	2.8e+07	1.834	0.071	-4.45e+06	1.07e+08
total_compensation	-8.695e+07	3.57e+07	-2.435	0.017	-1.58e+08	-1.58e+07

Omnibus:	26.278	Durbin-Watson:	1.453
Prob(Omnibus):	0.000	Jarque-Bera (JB):	117.633
Skew:	0.732	Prob(JB):	2.86e-26
Kurtosis:	8.472	Cond. No.	4.04e+19

The simple regression performs relatively well with an R-squared of 0.71. The unemployment rates over there years have been driven by separations which as expected have a positive correlation with benefits such as health insurance, paid vacation and total compensation. Interestingly enough, the industries are not significant when it comes to unemployment rates which is counter intuitive from what has been observed, especially during systemic events such as the COVID-19 pandemic. This indicates that assumptions and key inputs need to be revisited and ideally recursively improve the model. The data points are also sparse and there is a need for additional data points to include more different economic cycles.

NEXT STEPS & FUTURE WORK

Additional datasets

- We would want to go back further in time to observe other instances when employment was affected like the 2008 Recession or adujust the time frame analyzed by looking on an annual or quarterly basis
- Pull in finalized 2021 data since some datasets stopped in Q3 2021

Web scraping

- Repeat process with LinkedIn's API which:
  - Provides better sources of data because the output is more structured and
  - The platform is more centered around employment conversations from both an employee and employer perspective
- Since the ‘great resignation’ phenomenon was coined only 2 years ago, we would have to identify proxy/reference data for the previous years to do a historical analysis of employee sentiments.
- We can also repeat these processes with platforms such as Reddit or Facebook.

Prediction/optimization model

- Building and testing a better prediction/optimization model for HR teams that can be used to design targeted solutions for ideal work environments without compromising the bottom line

ACKNOWLEDGEMENTS

The findings and conclusions in this report are those of the members of Team 12, and do not necessarily reflect positions of CorrelationOne. We are grateful for the support we received from the Data Science for All program staff; our teaching assistant, Jamila Smith-Dell; our mentors, Rob Studt and Shannon Leber; and our colleagues, specifically John Montalbo, who allowed us to leverage their knowledge and expertise to further our analysis, pressure test our results, and refine our conclusions. We appreciate the knowledge imparted by DS4A instructors, without which we wouldn't have been able to execute complex analyses using Python and Tableau.

Lastly, we would like to thank the larger data science community, particularly Stack Overflow and other public forums, for openly making resources available to aid new entrants to the field.

REFERENCES

DS
4A

COHORT 3
TEAM 12

1. New York Times, Workers quitting their jobs hit a record in the U.S. in August, https://www.nytimes.com/2021/10/12/business/economy/workers-quitting-august.html, 2021.
2. CBC News, A record 4.4 million Americans quit their jobs in September, https://www.cbsnews.com/news/jolts-job-openings-quit-september-2021/, 2021.
3. SHRM, Why Are Workers Quitting Their Jobs in Record Numbers?, https://www.shrm.org/resourcesandtools/hr-topics/talent-acquisition/pages/workers-are-quitting-jobs-record-numbers.aspx, 2018.
4. Hollon, J., TLNT, New Research Shows Leaders Are Not Doing Enough to Develop TheirPeople, https://www.tlnt.com/new-research-shows-leaders-are-not-doing-enough-to-develop-their-people/, 2021.
5. Bureau of Labor Statistics, U.S. Department of Labor, Industry at a Glance, on the internet at https://www.bls.gov/iag/, (visited March 2022).
6. Twitter, www.twitter.com, (visited March 2022).
7. Backlinko, How Many People Use Twitter in 2022? [New Twitter Stats] https://backlinko.com/twitter-users, 2022.