Fluctuation of Employment in Tech Industry over Time

Yuang Hu, Wang Xiang, Shuchen Liu, Yiyan Sha

May 2023

1 Abstract

The tech industry has been a significant driver of economic growth in recent years, and the employment and stock price trends in this sector have been of great interest to researchers and investors alike. In this paper, we present a time series analysis of employment and stock price data in the tech industry, with a focus on understanding the impact of pandemics on these variables.

Our analysis is based on a large dataset that includes historical employment and stock price data for several tech companies. We use a combination of Spark, MapReduce, and Hive to process and analyze this data, leveraging the power of big data technologies to gain insights that would be impossible to obtain using traditional analytical methods. Our results show that there is a strong correlation between employment and stock price in the tech industry, with increases in employment generally leading to increases in stock price. We also find that the COVID-19 pandemic had a significant impact on both variables, with many companies experiencing sharp declines in employment and stock price in the post-pandemic age.

Our findings have important implications for investors, policymakers, and industry leaders, highlighting the need to monitor employment and stock price trends closely in order to stay ahead of market fluctuations and capitalize on emerging opportunities.

Contents

2 Introduction

“Pandemics create severe disruptions to a functioning society.” (Antipova 2). More specifically, the COVID-19 pandemic has had far-reaching impacts on economies worldwide, with various industries facing unprecedented challenges. The multifaceted economic and social aspects intersect in complex ways and are influenced inequivalently. The technology sector, being a significant driver of financial growth and innovation, is of particular interest to policymakers, investors, and researchers alike. In this study, we explore the relationship between stock prices and total employment in the tech industry across three major technology hubs: New York, California, and Washington. By leveraging big data tools and techniques, our investigation aims to provide valuable insights into the dynamics of the technology sector during this challenging period.

Prior to this study, we knew that the technology industry had experienced unexpected fluctuations in stock prices and employment during the pandemic, unparalleled with other sectors. According to the according to the Bureau of Labor Statistics (BLS), “While jobs took a hit during the COVID-19 pandemic with 17.6 million fewer positions in May 2020 than there were in May 2019, tech occupations grew by 1.2 percent in that same period”. (Nalea Ko) However, a comprehensive analysis quantitively and qualitatively of COVID-19 Impact on Tech Hiring as well as the overall health of the industry remained scarce.

To advance our knowledge on this subject, we are motivated to harness the power of a combination of big data tools, MapReduce, Hive, Spark, along with Java programming to analyze time series data of stock prices and employment rates. This study is then unfolded as follows with those powerful tools: we initiate by offering an in-depth account of the data collection process, emphasizing data reliability and precision. Subsequently, we tackle the data preprocessing stages, encompassing cleaning and normalization procedures, which are indispensable for readying the data for ensuing analysis and comparisons. Thereafter, we immerse ourselves in the exploration of machine learning techniques designed to discern trends and patterns within the time series data like multidimension regression. Lastly, we unveil our findings and engage in a thoughtful discourse on the ramifications for the technology industry, investors, and policy architects. The elaborate workflow is clearly demonstrated by figure 1 and will be further explained in Part 6.

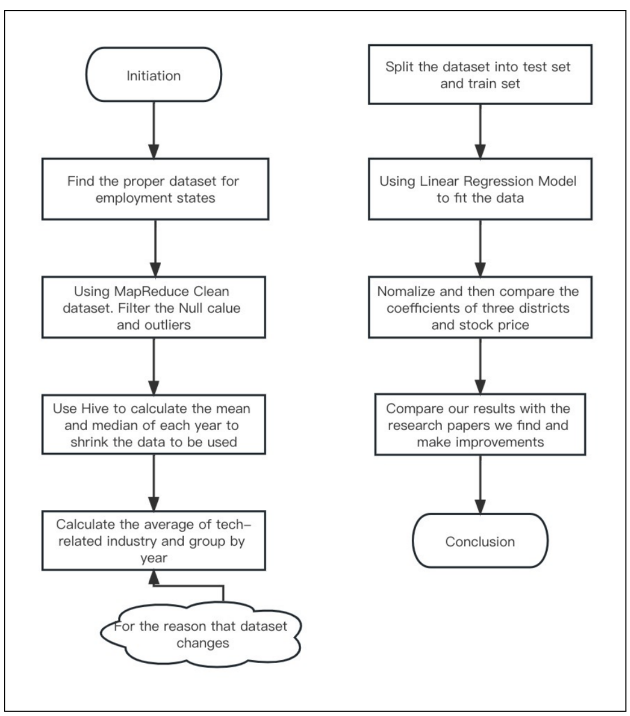

Figure 1: Diagram

3 Motivation

The COVID-19 pandemic has had a significant impact on the global economy, with many industries facing widespread layoffs and financial instability. Among the most affected industries is the tech sector, which has experienced significant disruption due to the pandemic's effects on supply chains, global demand, and changing consumer behaviors. As someone who wants to pursue a career path in this industry, we wanted to examine industry insights from this research to be better aware of the challenges and uncertainties facing this sector in the post-pandemic age.

Through this research, we aimed to shed light on the unique challenges and opportunities facing the tech industry in the post-pandemic age and to provide a data-driven analysis of the relationship between employment and stock prices in this critical sector. Ultimately, we hope that this research can help inform and guide industry stakeholders, policymakers, and aspiring professionals in navigating the complex and rapidly changing landscape of the tech industry.

4 Related Work

4.1 Related Work Description 1

New York City Department of City Planning and, Housing, Economic and Infrastructure Planning Division examine that: the comparison also reveals that, between the third quarters of 2010 and 2015, New York City's total average private employment from the Current Employment Statistics data increased by 15.2 percent, outpacing the national employment gains of 11.2 percent in the United States. Also, this report shows that, in the past five years, this sector has experienced consistent growth, which corresponds with the rising number of highly educated residents and the expansion of the tech industry.

It also shows that 'New York City's job gains since 2010 represent a remarkable recovery from the 2008 Financial Crisis'(Report P13). Our results approximately coincides with this article that the employments recovers from 2008 Financial Crisis. New York City has bounced back from difficult economic periods throughout recent decades; however, its recovery from the most recent recession has been exceptionally swift compared to previous instances.

4.2 Related Work Description 2

Geremew et al. [Geremew (2018)] examines seasonal and business cycles of U.S. employment from the 1980s to the 2010s by state and year, from which we can recognize tech industry has relatively high cyclicity when compared to other industries. Also, this article presents a general image of U.S employment conditions for our target years, industry, and cyclicity. It also examines the years after 2000 and around 2008 for fluctuations.

Roberts [Roberts (2013)] depicts and analyzes what happened to the economy, the job market of the tech industry, and stakeholders before and after the dot-cum bubble. Our result is generally consistent with this article that the bubble created huge opportunities but led to the collapse of job market after bubble burst.

4.3 Related Work Description 3

Doctor Jordi Gali, after conducting a thorough research and applying numerous economic models on technology sector fluctuation, concludes in The American Economic Review that "Yet, and to the extent that technology shocks are a significant source of fluctuations in those variables, we would expect RBC models to provide at least an accurate description of the economy's response to such shocks. For the majority of the G7 countries, however, the estimates of the effects of technology shocks yield a picture which is hard to reconcile with the predictions of those models". (Jordi 271) While his research was before the Covid 19, the result of our project further validates the conclusion as the machine learning model gives a poor correlation index for the fluctuation of the tech sector employment.

4.4 Related Work Description 4

Organizational Climate, Opportunities, Challenges and Psychological Wellbeing of the Remote Working Employees during COVID-19 Pandemic: A General Linear Model Approach with Reference to Information Technology Industry in Hyderabad.

This study focuses on the information technology industry in Hyderabad and provides valuable insights into the impact of remote work on employees' mental health and job satisfaction. It describes the main challenges faced by remote working employees during COVID-19 pandemic such as communication problems due to internet glitches and delays in communicating decisions. This has piqued our interest in the role played by the technology industry in the pandemic, which results in our decision of researching on the employment of tech industry employment dynamics and potential insights of this sector.

5 Description of Data sets

5.1 Data Set 1

One of the data set is the Industry Employment in New York City, which includes employment statistics by industry and time in NYC, giving a rough estimate of the types of jobs and corresponding number of jobs in the area. As the data set is updated by the Department of Labor, I recognize that this data set is convincing. In addition: It is considered that this could serve as a thorough and suitable source of information for the project because it includes monthly data from 1990 to 2022.

There are 18 columns in the data set: Area, Year, Series code, Area Name, Industry Title, Monthly data(in 12 months), Annual. Area is a string column, which refers to New York City and is all the same for this data set. Year is an int column that gives the year of the employment record, formatted as yyyy. Series code is a big int column that gives the corresponding numerical code of the industry title. Area name is a string column that is still the same for every row of data. Column 6 to Column 17 are 12 monthly data about employment condition. The last columns, which is Annual employment data, gives the average of that year.

5.2 Data Set 2

One of the dataset is the Industry Employment in California Counties, which includes employment statistics by industry and time in all California counties, generally to reflect an estimate of the number of jobs in the area. The dataset is credible as it can be downloaded from the State of California Employment Development Department, and it is derived from Current Employment Statistics (CES) dataset. In addition, as the dataset includes time period from 1990 - Present (2023) and is updated monthly with over 800k records, it is believed that it can be a comprehensive and appropriate source of information for the project.

There are 8 columns in the dataset: area type, area name, year, month, series code, industry title, seasonally adjusted, current employment. Area type is a string column, which describes the unit of analysis of each row, and this column is supposed to be all "County" in this dataset. Area name is a string column describing the name of each county of the row. Year is an integer column that represents the year of the employment record, formatted as yyyy. Month is a string column that represents the month of the employment record with full month date format. Series code is an integer column that describes the code that identifies the specific series of industry of the record, and industry title is a string column that describes the industry name, corresponding to the series code column. Seasonal adjusted is a boolean column that describes if the record includes seasonal adjustment, with N being No and Y being Yes and mostly N in this dataset. Current employment is an integer column that describes the number of jobs in this record.

5.3 Data Set 3

Name: Historical current Employment data in Washington State and labor market areas, from 1990 to 2022, not seasonally adjusted; Description: Overall, the data is split according to different counties of Washington State, each keeping track of the monthly non-agricultural employment of different industries from 1990 to 2022. Industry title: String- title of the section the specific industry belongs to of that specific row. Area Name: String- describe the official name of the geographical area of Washington State. Year: Integer- describe the year of the employment data as YYYY. Month: String - describe the month of the employment data as Text Current Employment: Integer- describe the total employments of the specific industrial section in each month.

5.4 Data Set 4

Stock Price Dataset: This dataset contains historical stock prices for top technology companies including Microsoft, Amazon, and Meta. The data covers a period of several years, starting from 2013 to 2021. Each row in the dataset corresponds to a specific day, and includes information such as the date, the opening and closing prices, and the volume of shares traded.

Employment Statistics (CES): This dataset contains employment statistics for the technology industry in California. it includes columns area type, area name, year, month, date, series code, industry title, seasonlly adjusted and current employment.

Tech Index Dataset: This dataset contains information on the overall performance of technology stocks on the Nasdaq exchange. The data covers a period of several years, starting from 2013 to 2021. Each row in the dataset corresponds to a specific day, and includes information such as the date, the opening and closing prices, and the volume of shares traded.

6 Analytic Stages

6.1 Ingestion Process

To analyze the relationship between stock prices and employment in the tech industry, we used MapReduce and Spark to perform data cleaning, profiling, and statistical analysis. To gain a more comprehensive understanding of the relationship between stock prices and employment in the tech industry, we conducted an in-depth analysis of data on the top leading companies in the industry, including Microsoft, Amazon, and Meta, as well as the Nasdaq Index on Tech stock prices. Our approach enabled us to investigate the overall trends and patterns in the sector by examining both the performance of individual companies and the broader market.

Our analysis revealed a significant increase in the tech index and historical stock prices for the companies under consideration during the pandemic. This finding is consistent with prior research showing that companies operating in the technology sector benefited from the shift toward remote work and the increased demand for online services during the pandemic. Notably, the rise in stock prices for these companies outpaced the growth of the broader market, indicating that investors viewed the technology sector as a particularly attractive investment opportunity during this period.

To further explore the relationship between stock prices and employment in the tech industry, we analyzed data on the employment rate in the "Professional, Scientific, and Technical Services" industry, which is where most technology-related jobs are classified. Our analysis showed that the employment rate in this industry fluctuated during the pandemic. This implies factors that may have contributed to this incremment, including increased demand for technology services that encourage companies to actively hire to provide and maintain their service.

6.2 Data Cleaning and Profiling Description

In the analytics stage of our project, we used MapReduce to clean and profile our data. The first step in this process was to remove any duplicate entries and unnecessary characters in the datasets that are not needed for numerical analysis. In the case of the stock price dataset, we did not encounter many outliers since the nature of the dataset is such that each date maps to the stock performance of that day. However, we still needed to remove the extra dollar sign in front of the stock price to deliver numbers that are suitable for further analysis.

Next, we conducted a data profiling exercise to gain a better understanding of the structure and content of our datasets. Specifically, we examined the size of the datasets, the number of rows and columns, and the data types of each variable. We also looked at basic descriptive statistics such as the minimum, maximum, mean, and median values of each variable.

For the employment dataset, we first examined the columns and filtered out the related industry of our interests by leaving only data rows in "Professional, Scientific, and Technical Services". We found that there were several missing values that needed to be addressed before we could conduct any meaningful analysis. We used MapReduce to fill in these missing values based on historical trends and patterns in the data.

Overall, our data cleaning and profiling exercise allowed us to gain a better understanding of the quality and completeness of our datasets, as well as the potential challenges and limitations of the data. By identifying and addressing missing values, removing unnecessary characters, and ensuring data consistency, we were able to prepare the data for further analysis and generate more accurate insights into the employment and stock price trends in the tech industry.

6.3 Analytics(Linear Regression)

6.3.1 Theory Base

Years are indexed as $t = 1, \cdots, T$; We denote $g(z_t;\theta)$ as the predicted value by the model, and r_t as the actual value recorded by the data set.
$g^{\star}(\cdot)$ can be approximated by a linear function of the raw predictor variables and the parameter θ:

$$g(z_t;\theta) = z_t^{'}\theta$$

Baseline estimation of the simple linear model:

$$\mathcal{L}(\theta) = \frac{1}{T}\sum_{t=1}^{T}(r_t - g(z_t;\theta))^2$$

Without additional command, Scala may use Robust objective functions using weighted least squares if the simple linear model doesn't work well:

$$\mathcal{L}_W(\theta) = \frac{1}{T}\sum_{t=1}^{T}w_t(r_t - g(z_t;\theta))^2$$

Specifically, Huber robust objective function:

$$\mathcal{L}_H(\theta) = \frac{1}{T}\sum_{t=1}^{T}H(r_t - g(z_t;\theta),\xi)$$

where

$$H(x;\xi) = \begin{cases} x^2 & if|x| \leq \xi \\ 2\xi|x| - \xi^2 & if|x| > \xi \end{cases}$$

Further more, the indicator R^2 is derived from Loss function $\mathcal{L}(\theta)$:

$$R^2 = 1 - \frac{T\mathcal{L}(\theta)}{\sum_{t=1}^{T}(r_t - \bar{r})}$$

6.3.2 Code Realization

```scala
import org.apache.spark.ml.regression.LinearRegression
import org.apache.spark.ml.feature.VectorAssembler
import org.apache.spark.sql.SparkSession
import org.apache.spark.sql.functions._
import org.apache.spark.ml.evaluation.RegressionEvaluator




// Load training data
val training = spark.read.option("header", "true").csv("final_code/Average_476.csv")

val training1 = training.withColumn("emp_n", col("avg_employment").cast("double")).select("year", "emp_n")

val training2 = training1.withColumn("year_int", col("year").cast("int")).select("year_int", "emp_n")

val assembler = new VectorAssembler().setInputCols(Array("year_int")).setOutputCol("new_col")

val spark = SparkSession.builder.appName("LinearRegression").getOrCreate()

val assembler = new VectorAssembler().setInputCols(Array("year_int")).setOutputCol("new_col")

val training3 = assembler.transform(training2)

val Array(train, test) = training3.randomSplit(Array(0.7, 0.3), seed = 1234)

val lr_model = new LinearRegression().setLabelCol("emp_n").setFeaturesCol("new_col")

val eval = new RegressionEvaluator().setLabelCol("emp_n").setPredictionCol("y_pred").setMetricName("rmse")

val pred = lr_model1.transform(test)

val training_summary = lr_model1.summary

println(s"RMSE: ${training_summary.rootMeanSquaredError}")

println(s"r2: ${training_summary.r2}")


#Idea get from https://spark.apache.org/docs/latest/ml-classification-regression.html#linear-regression
#Usage of functions helped by ChatGpt

```

Figure 2: Linear Regression Realization

7 Graphs

Visualization facilitates a better understanding of insights into employment and stock price changes.

Figure 3 illustrates the average employment in different areas over time on log scale. It can be observed that for California and NYC, the average employment fluctuates over time but has a general trend of increase. While for Washington state, the average employment in the tech industry increased from 1995 to 2000 but began to decline from 2000 till now.

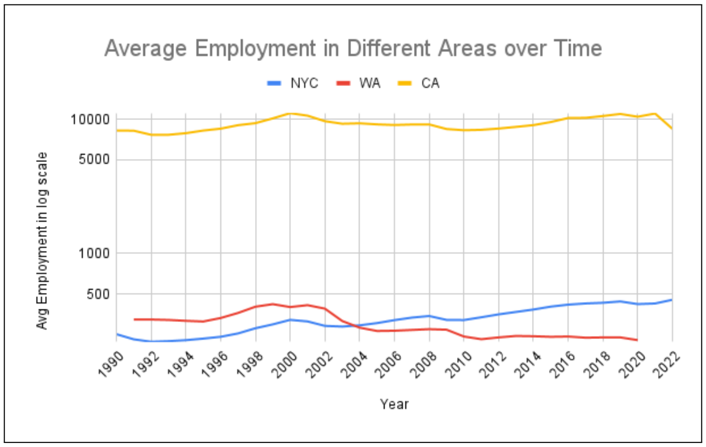

Figure 3

Figure 4 illustrates the change in employment for 2 industries that constitute the tech industry in Washington in our dataset. The fluctuations during other years are similar, while during 2001-2004, there are dramatic decrease, increase, then decrease especially in Computer system design and related services, and similar trends during 2017-2021.

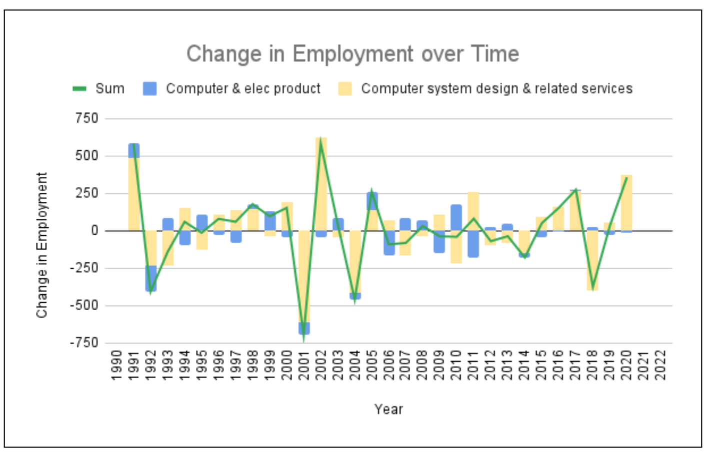

Figure 4

Figure 5&6 are two graphs presenting the stock prices of leading tech companies and the tech index during the recent 10 years. We can observe a general trend of increase over time and a prominent skyrocket in 2021, then a slight decrease from 2022.

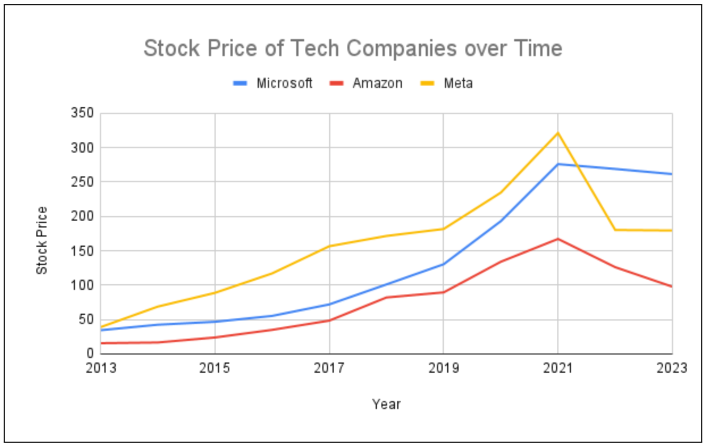

Figure 5

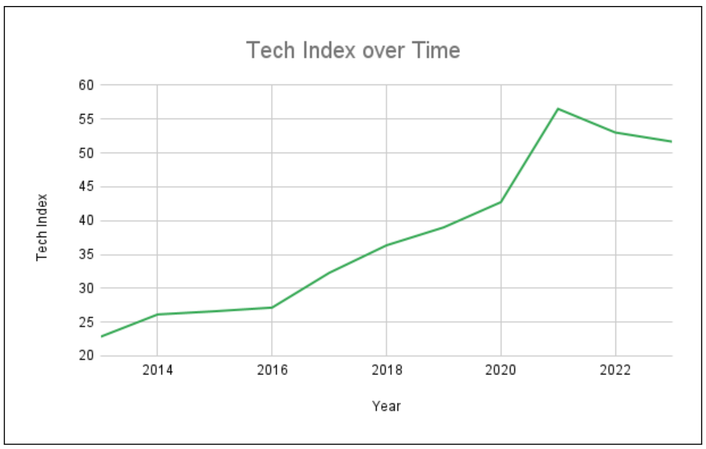

Figure 6

Except for the steady increase that is consistent with “professional services and information sector has grown steadily, in line with increases in highly educated residents and the growth of the tech industry (2010-2015)” (New York City Department of City Planning, 2016). There are 3 crucial time periods with abnormal fluctuations worth noticing, which are around 2000, around 2008, and around 2020.

Firstly, we can observe a rapid increase before 2000 and a plummet afterward in Figure 3, which is the result of the dot-com bubble and ensuing economic recession. With the advent of the web, investors are overly optimistic about internet-based companies so they contributed large investments. However, by 2000, many companies failed, and the bubble burst fueled the significant decline in job opportunities, as “students began to move away from computer science, which led in turn to a multiyear decline” (Roberts, 2003).

The great recession that happened around 2008 also influenced the job market remarkably. Both 3 lines in Figure 3 slope downward and changes in employment in Figure 4 are all below 0 around that time, indicating a huge decline in employment.

The last one, around 2020, is how the COVID-19 pandemic impacted the job market. All 4 figures betray similar trends of huge increase and then decrease after 2020. The unemployment resulted from the pandemic recovered due to the drastic increasing demand for online services including changes in customer behavior(online shopping) and working patterns (remote working). The employment status was stimulated in 2021 but decreased in 2022 partially due to oversaturated job market.

8 Conclusion

In conclusion, this study has provided valuable insights into the dynamics of the technology sector during the COVID-19 pandemic. Our analysis has revealed several key findings and contributions.

Firstly, our regression model has demonstrated that the high demand for technology during the pandemic has fueled the growth of e-commerce, as consumers increasingly relied on digital platforms for their needs. This growth has significantly contributed to the resilience and adaptability of the tech industry amid the unprecedented challenges posed by the pandemic.

Secondly, the time series analysis of tech index fluctuations and employment rates revealed that these variables do not follow a distinct pattern or function. This observation underscores the complex nature of the technology sector's response to the pandemic and highlights the importance of further investigation into the factors influencing these trends.

Moreover, our study emphasizes the significance of our methodological approach, including normalization, cleaning, and machine learning techniques, in uncovering novel insights and relationships within the data. By appropriately employing these methodologies, we have been able to derive meaningful conclusions with broader implications for the whole technology industry and intricated impact on people's daily lives.

Finally, our research has opened up new avenues for thinking about the sector-specific analysis and potential directions for future research. By building upon the insights and methodologies presented in this study, future researchers can continue to explore the myriad ways in which the tech industry has adapted and evolved in response to the COVID-19 pandemic, like considering the role of government policies and interventions, exploring alternative data sources, and applying cutting-edge machine learning and artificial intelligence algorithms for better prediction and comprehension. Those all ultimately expand to a full-scaled understanding of the sector's resilience and prospects for future growth.

9 References and Thanks

Thanks to Professor Malavet for her help during the process.
Refer to:
1. https://www.nyc.gov/assets/planning/download/pdf/planning-level/housing-economy/employment-patterns-nyc.pdf
2. Antipova, Anzhelika. "Analysis of the COVID-19 impacts on employment and unemployment across the multi-dimensional social disadvantaged areas." Social sciences & humanities open vol. 4,1 (2021): 100224.doi:10.1016/j.ssaho.2021.100224
3. Ko, Nalea. COVID-19 Impact on Tech Hiring Trends, April 21,2023
https://www.bestcolleges.com/bootcamps/tracks/covid-impact-on-tech-hiring-trends/
4. Gali, Jordi. Technology, Employment, and the Business Cycle: Do Technology Shocks Explain Aggregate Fluctuations? Source: The American Economic Review, Mar., 1999, Vol. 89, No. 1 (Mar., 1999), pp. 249-271
5. https://papers.ssrn.com/sol3/papers.cfm?abstractid=3599799
6. https://data.edd.ca.gov/Industry-Information-/Industry-Employment-in-California-Counties/nt76-4rha
7. Geremew, M., & Gourio, F. (2018). Seasonal and Business Cycles of U.S. Employment Federal Reserve Bank of Chicago. https://www.chicagofed.org/publications/economic-perspectives/2018/3
8. Roberts, E. (2003). What happened during the downturn in the 2000s?
https://cs.stanford.edu/people/eroberts/CSCapacity/CSCapacity-History2000s.html