

# 1 Abstract

2022 hasn't been the most kind year to us in the tech domain. In general, world economy has been rapidly declining and showing negative signs in every direction. Extremely high energy prices, companies scoring all time lows in the stock exchange, record inflation rates, are all negative signs of the future. As a result, many companies were faced with the challenge of losing many sources of future finance and the immediate need to halt their growth plans and sustain. This has put a huge pressure on many companies that were already massively expanding and were not prepared for this. In this project, we try to build a data set of economic factors that might be putting pressure on both private and public companies to shrink and layoff employees. Once we've acquired said data, we'll try to quantify the effect of each factor on the magnitude of layoffs and see if our exploration will show us which factors were the most pressuring on companies and might have caused the loss of jobs.

# 2 Introduction

Periods of recession often follow periods of high inflation rates or extreme events, like wars, bursting of the house bubble, or stock market crashing. In 2019, the COVID pandemic has started and within a few months forced entire countries into lock downs. This sacrifice entailed a bleeding economy and incoming shrinkage. However, things returned back to normal in 2021 with companies scoring huge profits and their stocks reaching all time highs. But that didn't slow inflation down, and with the Russian invasion of Ukraine, economy couldn't take it any more. The quick hit accompanied by the high energy price and rising FED and ECB rates placed increasing and crushing pressure on many companies. Companies were already in a growth stage following the fast recovery after the lock downs in 2021. This rapid change of temperature caused a huge shock that businesses had to absorb and re-plan their future. Unfortunately, one of the first techniques that companies follow to mitigate the increasing pressure of a shrinking economy is to lower costs, which entails decreasing work force.

Lay offs are extremely unfortunate. People lose their jobs when they least expect it. They still have to pay mortgage or are still on an H1-b visa where they have to remain sponsored. Losing your job can be one of the worst experiences possible. That's why it's really important to understand what kind of pressure would push a company into contraction and laying off. This way we can further predict if a company is at risk of laying off employees or help decision makers understand the true expected extent of their decisions.

The goal of this analysis is to: 1. extract sufficient data relating lay off events to economic factors at that time. 2. Using Bayesian Analysis, weigh the effect of each factor on these events and try to find factors that are directly responsible for lay offs (However, this analysis is only going to find correlation and not true causation).

Exploratory data analysis to understand the diversity of lay offs and which industries were

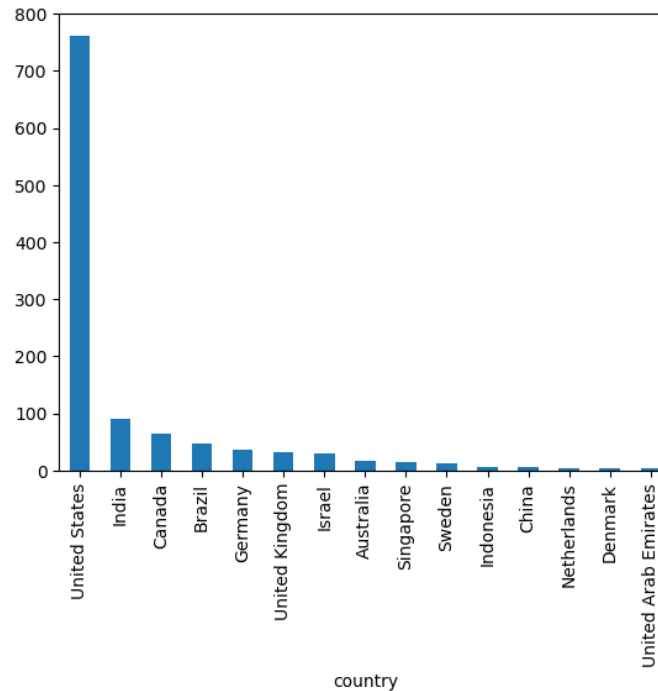


Figure 1: Number of Lay off events by country

most affected was performed in the notebook **Data Extraction**. We can summarize the EDA with figures: 1, 2, 3, 4, 5.

### 3 Building a Data set

Our goal is to attribute economic factors to lay off events. In order to do this we need to have a data set that includes as much lay off events as possible, while also having an economic snapshot at that time of both the world economy and the company's economics. Building such data set is not hard given that the internet has almost everything we need. Public information includes both information about companies and economy on a daily or even hourly basis.

Starting off with lay off events data, there has been great effort to keep track of every lay off event in the tech domain since the inception of the pandemic. This tracker has been made public under the domain [Layoffs.fyi](https://layoffs.fyi) which included information from 2020 all the way until today. This data has been made available as a [Kaggle data set](#) for analysis and research. This data set includes for each lay off event: the company's name, location, industry, funding stage, number of employees laid off, percentage of employees laid off, and the date of the event. Because we have the date, we can actually search for economic data on that date and have extra features that resemble a snapshot of the economy at that period.

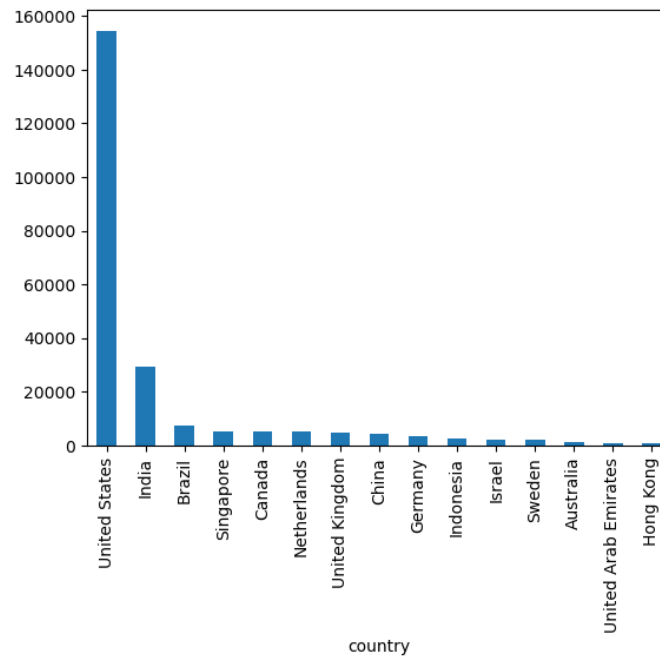


Figure 2: Total number of employees laid off by country

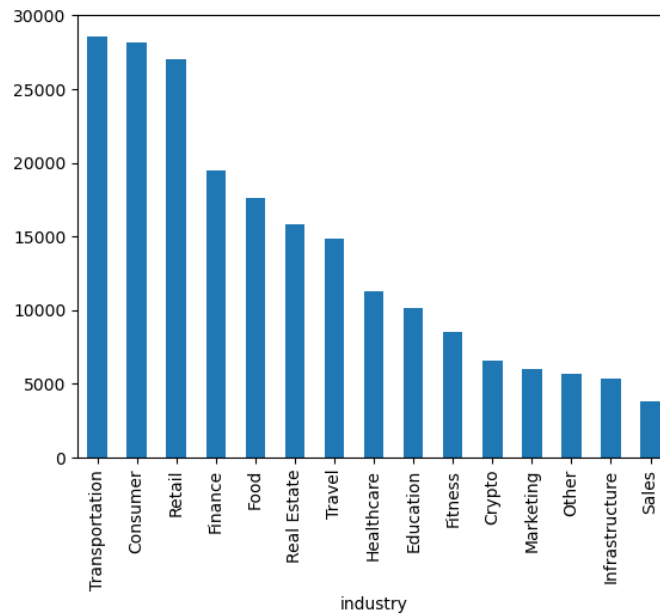


Figure 3: Total number of employees laid off by industry

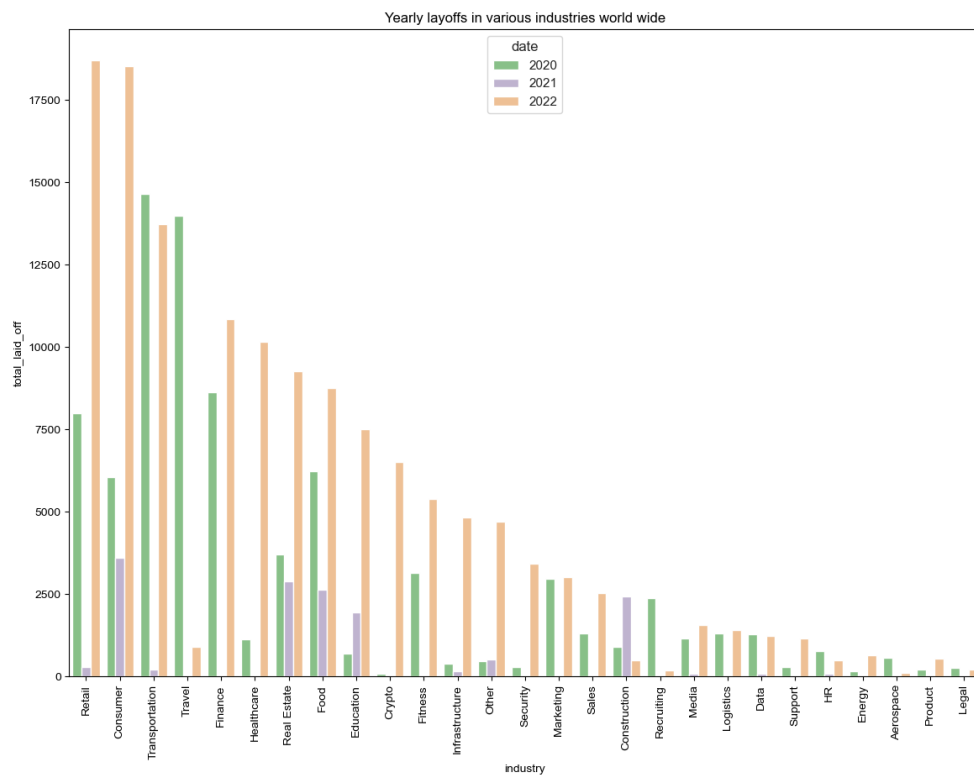


Figure 4: Yearly total number of employees laid off by industry

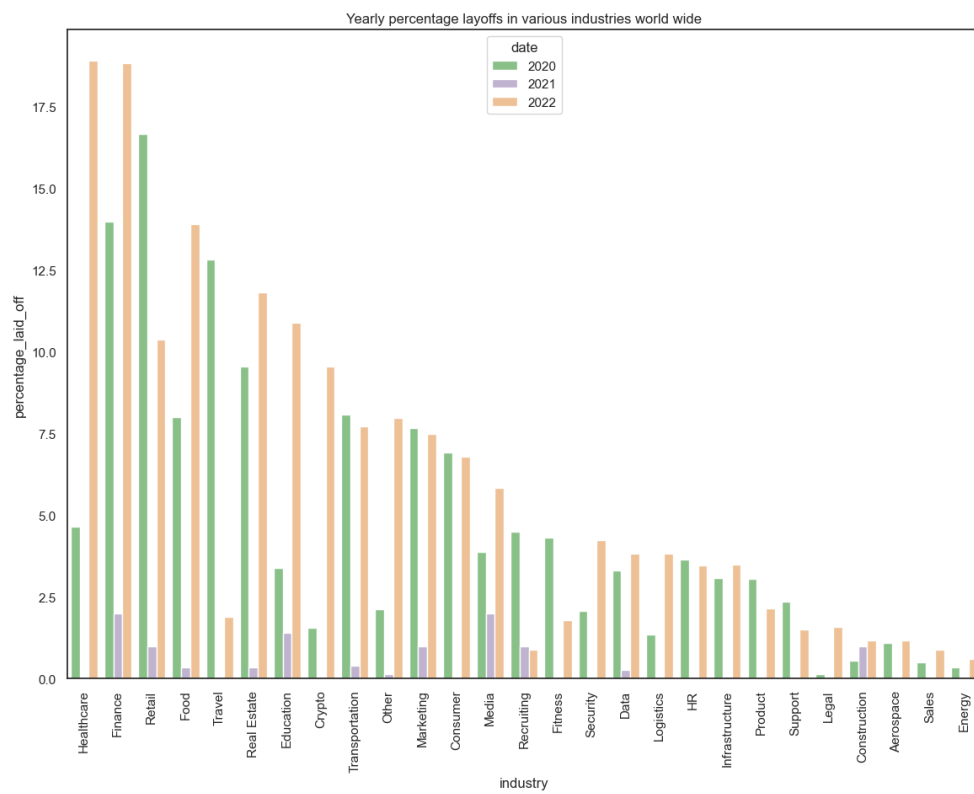


Figure 5: Yearly percentage of employees laid off by industry

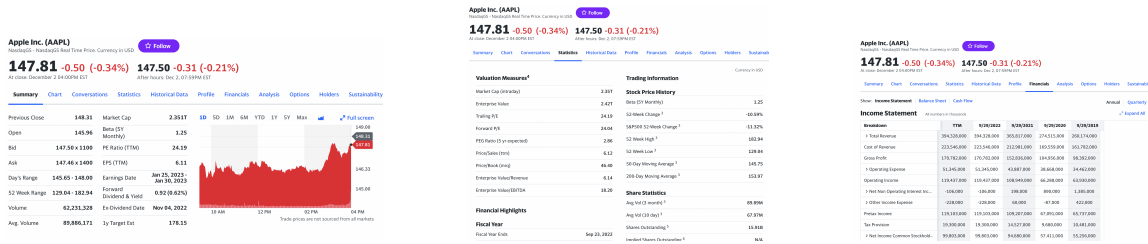


Figure 6: Public Data on Yahoo Finance

A good strategy for gathering economic data is to pivot on the date of the event and take a snapshot of the economy. However, we know that these factors aren't really born in the moment, they're cultivation of months of declining. Instead of taking a static snapshot at that time, we can also consider time factors like: decline over the past month, 6 months, and 12 months. This way, we can capture the down trends of certain trend based factors.

There are many sources online for economic data, however, some are free and some are paid. Among those solutions, Yahoo Finance seems to be the most established and easy to use. Manually testing the search functionality showed that they have a lot of data on each company whether it's private or public. This is very important as a huge portion of lay offs happened at companies that are not public yet and it wouldn't be a good idea to throw them away.

Now the question is, which factors to include? For public companies, we have a lot of different data points that we can extract, however, for private companies, we're very limited. Due to this distinction, we should split our analysis and our data into 2 different subsets: public companies, and private companies. This way we don't have to limit ourselves to the least common factors between public and private. Instead, we can get as much data as possible and attribute the effects according to the type of company.

### 3.a Public Companies

The best part about public companies is that they're mandated to share a lot of information about their economics to the public. Thus, making it easier to look into their performance and understand what leads to a certain state. If we look at Apple for example, we can find summary data describing the performance of the company in the stock market. We can also find historical data, financials, and statistics. As we try not to be overwhelmed by the data, we have to pick certain factors that can be comparable across all companies. For example, we shouldn't use any absolute quantities, rather relative quantities are better for analysis. For example, instead of using total revenue, we can use Return on Assets and Return on Equity, which would be a percentage relative to the company's assets/equities. This way we don't skew our model by meaningless big numbers.

The first challenge we meet is to match a company's name to its stock market ticker. A stock

market ticker is an ID that references a certain company's profile in the stock exchange and can be used to pull the company's data from sources like Yahoo Finance. To overcome this challenge, we automate the task of searching for a company by name and matching the name to the ticker in the first retrieved result. Empirically, this works very well and according to manual inspection, returns the correct results.

After matching a company to a ticker, it's trivial to pull all available company data, but as mentioned before we want to only keep relative factors. We also don't want to drown our models with variables, instead, we want to focus on a select but representative set of factors. The list of factors that are extracted are:

- profitMargins
- revenueGrowth
- shortRatio
- forwardEps
- currentRatio
- earningsGrowth
- returnOnAssets
- payoutRatio
- operatingMargins
- beta
- fullTimeEmployees
- earningsQuarterlyGrowth
- returnOnEquity

The idea from this selection is to only use relative variables and to use variables representing the company's performance in the stock exchange as well as the company's financials. This process yield a data set of 470 public company with lay offs.

These factors will also be coupled with the performance of the S&P500 as lagged variables at the event date. We will use the change percentage of the S&P500 over the past 30, 180, 365 days.

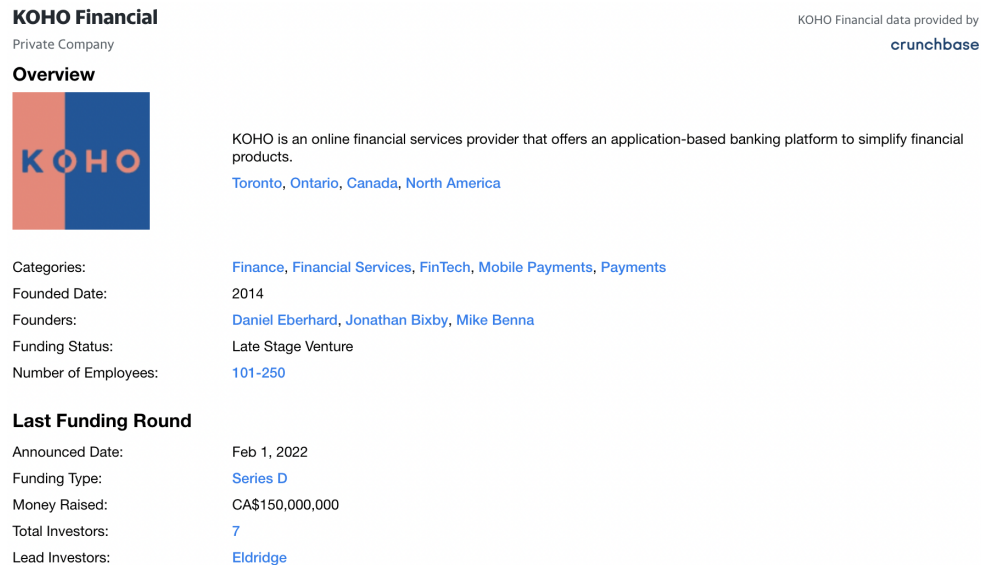


Figure 7: Private Company data available for KOHO Financial

### 3.b Private Companies

In case of private companies, we know these are not companies that have a defined ticker and are in the stock market, so what we can do is see if Yahoo Finance has a ticker for them (There are tickers for cryptocurrencies, for example), if they do then we pull everything and treat them as a public company. If they don't then we search Yahoo Finance for a private company's profile that matches this name and then we pull the corresponding data.

The problem with this approach is that the data we'll get is very limited. We don't have any factors describing performance, we don't have any factors describing financials, and we don't have any factors describing statistics. We're only left with some information about the industry, number of employees, and amount raised so far (Figure 7). That's why this part of the data set is troublesome and we'll see from the analysis if a model can in any way associate these features to the lay off events. This process yield a data set of 974 private companies with lay off data.

These factors will also be coupled with the performance of the S&P500 as lagged variables at the event date. We will use the change percentage of the S&P500 over the past 30, 180, 365 days.

### 3.c Negative Samples

So far we've only been pulling data for companies that witnessed lay off events, while this is the core of our interest, we still need to gather data of companies that did not have lay off events. This way we force our model to learn the distinction between companies that are



laying off employees versus those who are stable and can ride the wave.

In order to do this, we sample 2000 random tickers from NASDAQ and remove any intersection with our laying off companies then we extract the exact same features we extracted before. Yet, we assign 0 to the number of laid off employees and the percent of laid off employees to indicate that no lay offs happened.

On the other hand, for private companies, we built a list of unicorns (startups exceeding 1 billion dollars in valuation) and also a list of startups that aren't unicorns. The purpose was to diversify private companies across the money raised spectrum and to target both Early and Late stage ventures. The extracted set of 2297 companies was then subjected to the same process of feature extraction as positive samples.

As for the lack of date for the event on the negative samples, looking at layoff data it seemed that the mode of the dates was April 2020. Thus all negative samples were assigned temporal features from that period.

### **3.d Pre-processing**

There were 2 steps for pre-processing this data set. Due to the large number of categorical variables, we needed to have mappings that aggregate these categories to more general categories and thus have less variables in the model. For this, a mapping was manually constructed to convert 377 different categories into 27 general categories. This applied to the features: Industry, and Category. Another mapping of 33 different stages was constructed to map the Stage, Funding Type, and Funding Status features into 4 general values.

The second step was to transform those mapped categorical variables into one hot encoding variables indicating whether if a row in our data set belonged to a certain category.

## **4 Analysis**

Coming from a Machine Learning background where everything is a weight and non-linearity doesn't expose you to the beauty of the Bayesian methods. In Machine Learning, fitting a model is basically just finding the set of weights that minimizes an objective function. However, in Bayesian methods, treating each weight as a random variable is a really powerful method. The effect of that is that we can explore model uncertainty on the level of certain factors. Studying credibility sets and variance of each random variable helps understand and quantify how good your model is. If the model is really confident in fitting the observations, then we'll notice a narrow credibility set and lower variance. On the other hand, if we have a wide range and high variance, then we can expect our model to be uncertain in itself and in its ability to explain observations.

Based on this power of Bayesian methods, we will try to fit a logistic regression model where we assign a random variable to each feature in our data and try to predict a Bernoulli random

variable which is the percentage of workforce laid off (As per Figure 8). After sampling and observing our data, we'll inspect each random variable's posterior distribution looking for answers. If we see a variable whose posterior looks well defined with narrow credibility sets and low variance, then we can deduce that said factor associated with this random variable has a direct effect on the percentage of lay off.

However, if our posteriors do not look like they're well defined with low variance, then we can assume that this factor is not of importance. There are other methods to do this like measuring Deviance of the model or hypothesis testing, but the simplicity of observing posteriors and following intuition might be enough for this task.

## 4.a Public Companies

The leverage public companies have is that there are a lot more continuous features that track a lot of different factors like financials, statistics, and market performance. Before starting any analysis you can already expect a better fit for public data than private data.

To analyze our factors we fit a Bayesian Logistic Regression model on the percentage of employees laid off. The model specification is as illustrated in Figure 8. We define our weights as random variables sampled from a non-informative normal distribution prior defined as:  $\beta \sim \mathcal{N}(0, 1000)$ . As for missing values in the data, we mask them so that PyMC can handle them as missing data. An intercept is added to the original data by horizontally stacking a vector of ones so that we can have a bias term -which is  $\beta_0$ .

We run our MCMC sampling with a burn in of 5000 samples and we sample 5000 more samples for our posterior with a target acceptance rate of 95%.

The resulting posteriors (Figure 9) resemble normal distributions for most of our variables. Let's explore factors whose beta's credible set doesn't include 0.

- Revenue Growth, mean: -0.004
- Return on Equity, mean: 0.682
- S&P500 Change 30 days, mean: 13.210
- Industry: Crypto, mean: 22.991
- Industry: Finance, mean: 18.555
- Industry: Food, mean: 21.239
- Industry: Healthcare, mean: 17.681
- Industry: Infrastructure, mean: 21.575
- Industry: Product, mean: 21.289

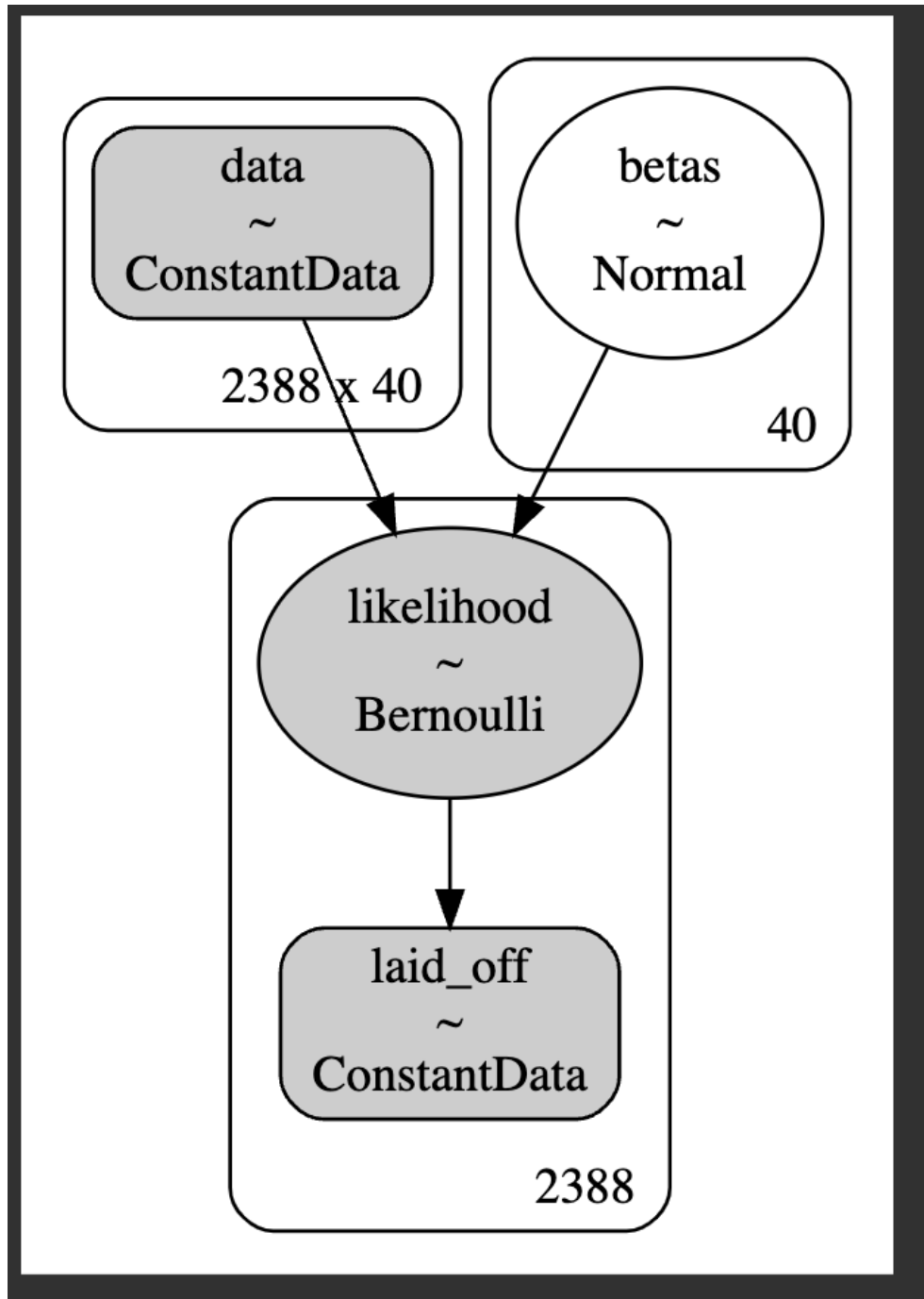


Figure 8: Bayesian Logistic Regression Model

- Industry: Real Estate, mean: 20.229
- Industry: Retail, mean: 20.460
- Industry: Transportation, mean: 21.085
- Industry: Travel, mean: 21.886

It looks like we have 3 factors affecting lay offs: Revenue Growth, Return on Equity, and the S&P recent trend. For Revenue Growth we see that the mean is small which means that this factor isn't that strong yet it is significant. The same can be said on Return on Equity, although it's orders of magnitude higher than Revenue Growth. But the highest mean can be seen associated with the S&P 30 days Change. Indicating that possibly short term high pressure from the stock market could have a butterfly effect on companies or could be a heuristic that investors are cashing out and a recession is coming.

On the other hand, we notice that some industries are more probable to be associated with lay offs at dire times. These industries are: Crypto, Finance, Food, Healthcare, Infrastructure, Product, Real Estate, Retail, Transportation, and Travel.

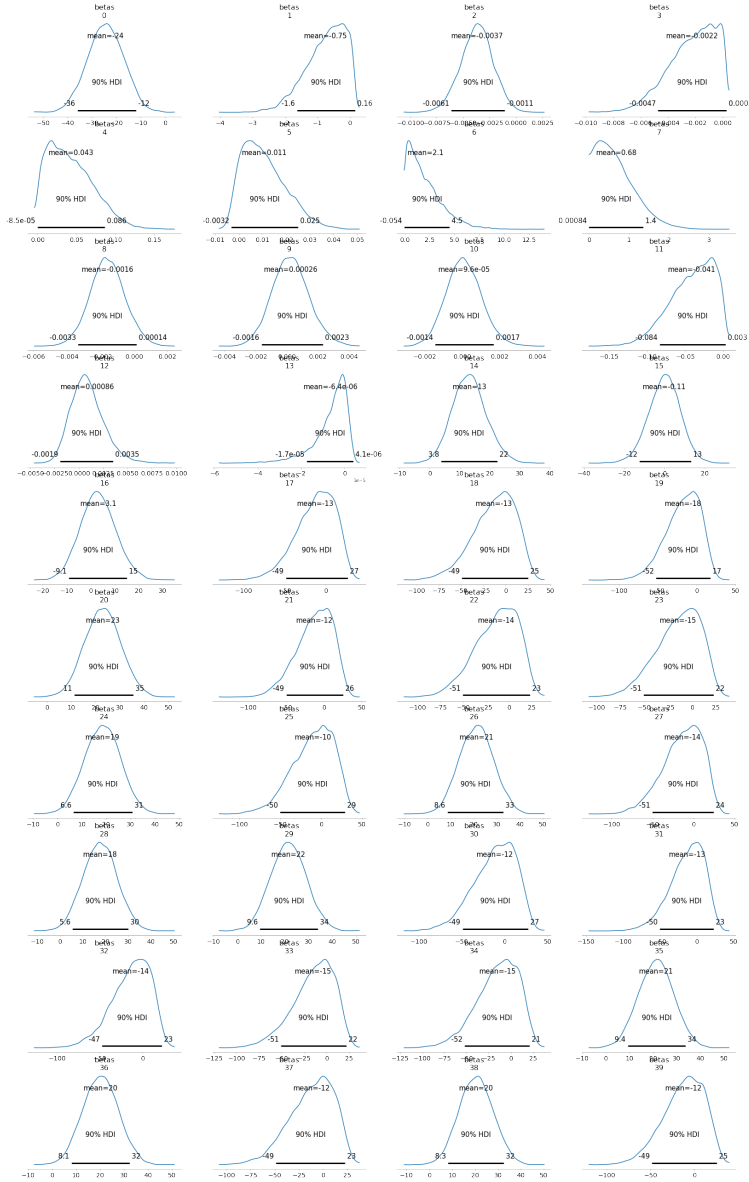
## 4.b Private Companies

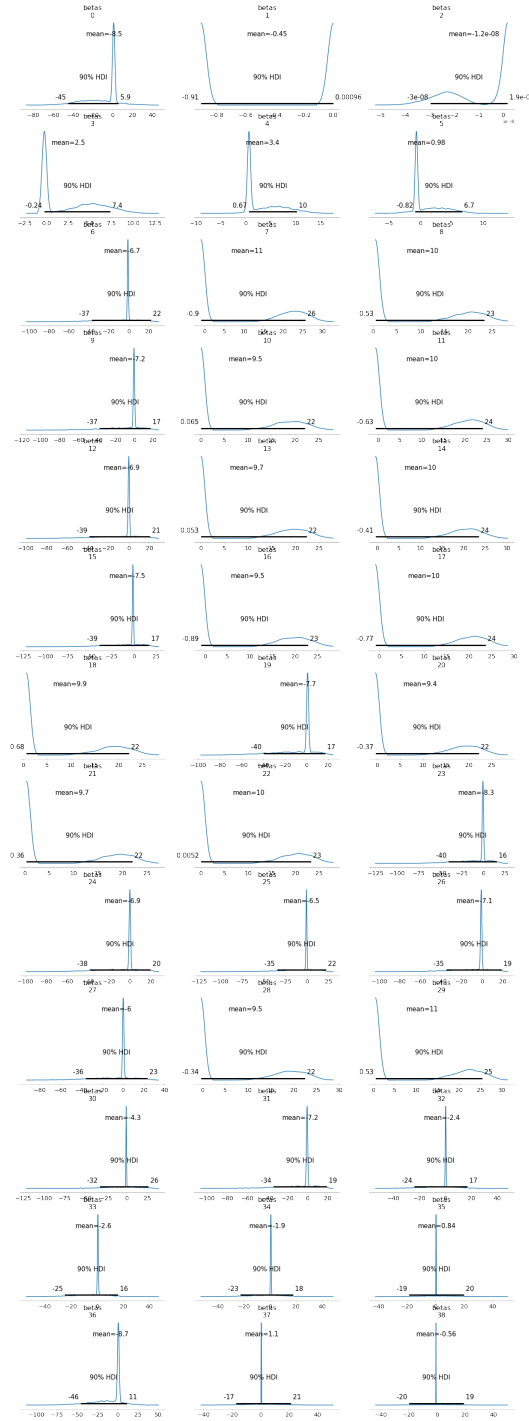
Following the same procedure to fit a Bayesian Logistic Regression model on Private companies didn't seem fruitful. The posterior distribution (Figure 10) does not seem well shaped and the values of Gelman-Rubin statistic  $\hat{R}$  is not close to 1 which indicates lack of convergence. Different trials were done to mitigate this by raising target acceptance rate or sub sampling the data, yet convergence didn't seem possible.

However, if we discard our doubt and look at what the model finds, we can see that the S&P500 Change over 6 months might be a significant factor. Also some categories showed a non zero credible set including: Data, Consumer, Finance, Infrastructure, Media, and Support. The full experiment along with the results can be found in the notebook "Bayesian Analysis"

## 5 Conclusion and Future Work

In this experiment, we built a data set of economical factors and associated it with available lay off event data in the hope of finding which factors affect lay offs the most. The result of this analysis was that for public companies we found that Revenue Growth, Return on Equity, and S&P short term change can be important factors affecting a company. We also noticed that some industries are more susceptible to lay offs and these industries are: Crypto, Finance, Food, Healthcare, Infrastructure, Product, Real Estate, Retail, Transportation, and Travel.

Figure 9: Posteriors of our  $\beta$  vector on Public companies

Figure 10: Posteriors of our  $\beta$  vector on Private companies

As for future work, the extent of economic factors on private companies was not assessed due to lack of convergence possibly due to lack of good features. Solving this problem would yield more interesting data. Other directions of research might be trying a pure binary classifier instead of predicting the percentage of laid off employees. Construction of negative samples is always tricky, researching a better alternative to sample negative samples might help the analysis. As for the analysis itself, the possibility of treating this analysis as testing two hypothesis: Whether a certain factor affects a company's lay off decisions or not is a possibility.