Kerry Effinger

Predicting Tech Industry Layoffs: A Regression Model

April 25, 2023

[Kseffinger04@gmail.com](mailto:Kseffinger04@gmail.com)

Executive Summary

With news of companies laying off employees in technology positions constantly appearing in headlines, it seems as if there is a never-ending trend of tech layoffs. To investigate the characteristics of companies engaging in these layoffs and the frequency with which they occur, data was collected on layoff events as well as information about the company that made the decision to let these people go. The attributes of these companies was studied to find patterns and to create a model that can predict whether a company in the technology industry will conduct a mass layoff and how large it will be. The major findings of this project were that location and business stage were important to layoffs. The cities with the largest layoffs were the San Francisco Bay Area, Seattle, Austin, and New York City, and Businesses in the IPO stage (publicly traded) had the most layoffs. The major conclusion from the linear regression model was that a linear regression model would not be appropriate in determining the size of a layoff event, and a different model should be used for predictions.

Introduction

The website, [layoffs.fyi](https://layoffs.fyi/), is a website created by startup founder, Roger Lee, to track companies that have laid off employees (in small numbers or very large) since the beginning of the COVID-19 pandemic in March of 2020. While the layoffs in the past several months are likely not a result of the initial shocks of the pandemic, the website still tracks the layoffs on a regular basis, with new entries nearly every day. For this project, only layoffs since the beginning of 2023 are considered and continually added to the dataset as more data becomes available. The goal of this project was to gain insights into the layoffs that are occurring in the United States to see if there are any patterns. Additionally, a predictive model that can predict how many people a company is expected to layoff was the ideal outcome. Other objectives of the project were to design an informative and easy-to-read dashboard that summarizes the work from the semester and also to expand my skills in database management, data analysis, and dashboard design.

Project Details

The first step of the project was collecting data from layoffs.fyi. Web-scraping was thoroughly researched and attempted, but manually entering data into Excel from layoffs.fyi proved to be the best use of time in order to make progress in the project. After data was collected from layoffs.fyi, it was loaded to and stored in MySQL Workbench with some SQL queries. A connection was made directly from MySQL to a Jupyter Notebook for data analysis and to Tableau for data visualization. With the data linked from MySQL, data analysis and creating a regression model to predict future layoff events could begin. Jupyter Notebooks with the language Python were used for these steps. The linear regression model was created with python scripts, and though its results were not very accurate or helpful, they were still presented as what *not* to do if attempting this type of prediction. At a minimum, this method can be eliminated from consideration to save time in the future if attempting this project again. Finally, the software Tableau was used to create visual graphs and dashboards to summarize the results of the project. A visual heat map showing all of the layoffs throughout the country was very informative.

Reflection

The first success of my project after deciding my topic, was the discovery of layoffs.fyi. I came across the website accidentally when I was researching tech industry layoffs early in the semester. I discovered that all this data was conveniently collected and formatted, and I was able to base my variables for the remainder of the project off this data source. Next, once data could be collected, I think the exploratory data analysis went well. That was something we had experience with already, and it was encouraging to make progress after a few setbacks with the web scraping and database setup. I was able to discover patterns in the data and learn about my specific dataset to make some inferences about the state of the tech industry overall. Also, I was able to establish a connection between my SQL database and a python notebook with few issues. This only required a few lines of code, but this small piece really impressed me, and I see how it could be so valuable to future data analysis projects. That was probably my favorite technical skill I learned. Along with the exploratory data analysis, making the Tableau dashboard to visualize my data also went well. Fortunately, having a few previous projects with Tableau made it so I could figure out how to show everything I wanted to. I still looked specific things up and watched a few tutorials to refresh my memory, but I was pleasantly surprised that this step was not time-consuming. Next, I feel that designing and presenting my poster went well. There were many last-minute additions, but I was pleased with the final product. I felt I had a good grasp on the material, so presenting at the Celebration of Student Research and answering questions was not too intimidating. Further, I found a sense of accomplishment and relief presenting the project at this stage (though not perfect), and I really enjoyed seeing my classmates’ and my work come together.

I worked for a long time on trying to find a way to web scrape the website layoffs.fyi. Since the data was contained within an Airtable, and the author made it unable to be downloaded, this proved to be much more complex than any web scraping tutorials I read. While there still may be a way to accomplish this task, I decided in the interest of time to manually copy down the data from the website into an Excel spreadsheet. This became my data collection system, and though it was not ideal, it allowed me to progress in the project. After collecting data from layoffs.fyi, I wanted to collect financial data, such as revenue and earnings per share, to use as additional explanatory variables to predict layoffs from a company. Unfortunately, Yahoo Finance’s API is no longer available to easily extract data. However, there are many other packages that connect to Yahoo Finance, and I was able to download them and use them in Jupyter Notebooks. Once again, this did not work out, as several of the metrics were unavailable. Weeks later, I managed to access stock prices and could come up with a way to calculate a few financial ratios. However, only publicly traded companies with stock symbols could be accessed in Yahoo Finance, which was less than half of my nearly 400 rows of data. Further, an even smaller number of those publicly traded companies could connect directly to the Yahoo Finance package. Even if I spent the time writing the code to collect some financial data, it would only apply to a very small number of the rows, and the model may still not be accurate due to a small sample size. I decided to move on with the other descriptive variables I had, and that was when I encountered trouble with MySQL Workbench. At first, no queries would run and the software crashed every time I tried. I finally was able to delete the program and revert to an older version which did not crash anymore. Then, I could not get all of my data to upload. This ended up being an issue with the blank cells of my data, and I realized I needed to fill those is with 0 if numerical data or NA if text data. All of this cost me some time, as I tried to resolve these issues in the background while also moving forward with other parts of the project. Finally, my predictive model did not work well. I did not expect a linear regression to be a good predictor for this data, but I wanted to use it as a starting point. Unfortunately, due to time constraints I was only able to run that linear regression and present it on my poster. It is clearly not a great fit, but at least I eliminated that model from consideration, and I know that another model would work better.

Time proved to be a constraint throughout this project. If I could have started collecting data a month or more sooner, then I feel I could have had a more successful predictive model. For example, if I had data from even some of 2022, then I may have had enough to formulate a decent model with the limited financial data I could obtain. Further, when collecting data from layoffs.fyi, I filtered the data to include only those companies based in the United States. However, knowing what I know now about the Yahoo Finance data and the need for publicly traded companies, I could have added a filter on the website and only collected data on publicly traded companies. If I could have started collecting data from mid-2022, then I feel I could have used some financial data as explanatory variables which may be a strong indicator of layoffs. Going in a different direction, if I wanted to run a logistic regression, with the dependent variable being 1 if the company laid people off and 0 that it did not, then I should have collected data differently. I could have been collecting data on random companies with similar attributes as the layoffs.fyi dataset that did *not* lay off employees. This may have proved to be difficult as well since I would have to search for companies with this description and it would not all be in one place like layoffs.fyi. If I had thought of this sooner and had the time, I may have shifted my model.

My advice to others working on a similar project would be to think ahead and think about the bigger picture. I know it is difficult to not get caught up in the details and problems that arise at every step of the project, but forward thinking will really help. If you can think about the final product at every step, then maybe you can plan ahead to set yourself up for success. For example, if I had thought more about the final predictive model earlier in the semester, then I may have been able to collect the data I needed to make the model the best it could be. Additionally, start thinking of your topic idea well in advance. Fortunately, my first idea worked out for me, but I know sometimes it can take a while to come up with a viable project idea. Finally, I am not sure how feasible this is, but if a data collection system could be established before the start of the semester, that could save time and help students be successful. It does not seem like creating a data collection system alone should take up an entire semester-long class, but if I had had my data collection started in the fall semester, I could have spent more time on other areas of the project and also had more data to work with.

Although web-scraping did not work out for me, I still learned much about the process which was previously unfamiliar to me. I feel that in the future, with some patience and a slightly simpler website, I could achieve the feat of collecting data from web scraping. If nothing else, I am more familiar with automated data collection and more aware of its many applications. As for working with MySQL Workbench, I was definitely frustrated at several points. One solution was reverting my software back to a previous version, as the newest one had many bugs that impaired its functionality. I also learned about uploading data to the database, and how specific the formatting of the data source must be. These small tricks may come up again in future projects, and I am glad that I encountered them. If the exact same issues do not arise, I at least have the experience to think about issues differently and work to problem solve. Also, something I already was a aware of but certainly had to understand this semester, is that Mac laptops are often limited in this industry as compared to Windows or other systems. Having a Mac presented some challenges, and in my future career, I know now to choose a non-Mac system for myself when possible.

In conclusion, I wish my list of things that did not go well was not so long, but I certainly learned a lot during this semester, and that is what matters most. Like mentioned previously, I have learned about various techniques and tools that I did not have much knowledge or experience with before. I also have learned a great deal about problem solving in a technology environment (software can be very frustrating!). Since the processes are pretty much set up, I can continue to collect data and analyze it. If I have the time, I would love to come up with a better working model and see if I can actually predict layoffs with some accuracy, because it is a topic that remains important to me. Outside of this specific project, I would like to have my future work more fully automated, using web scraping if possible, and storing data to conveniently access at any time. I think automation will make things much more efficient, and knowing these skills makes me a more valuable employee. I feel confident that both the technical skills and soft skills that I learned this semester will be important to my future career, and I am grateful to have completed this project with some success.