Amariah Robinson
02/23/2025
DTSC 2301-001
Dr. Benedict and Dr. Blekking

<p style="text-align:center">Project 1 Portfolio</p>

Hello and welcome to my Project 1 Portfolio. 🙂

This portfolio is to highlight my insights in my research and to see my thought processes and analysis! The dataset that I will be using is from Kaggle, which is titled "Tech Hiring and Layoffs: Workforce Data (2000-2025). As you can see from the title, it is about various tech companies that have hired and laid-off employees in the last 25 years. So based on my dataset, I wanted to know if for the tech companies listed, if companies had more layoffs in the 2008 recession or during the COVID-19 pandemic, in the years 2020-2021. This question is interesting because it can show the effects of our economy and what it does to the everyday employee. This dataset is also useful because it shows time trends over the course of the last 25 years, which has had many technological advancements, controversially the biggest being the boom of Artificial Intelligence (AI). Many people in the tech industry would care to see what the job market is like and how it may or may not be improving, and can skew their decision on the job market and what company to even apply for. If a company has a small unemployment rate and has a huge amount of new hires, I would absolutely choose that over another company that is laying off thousands of people and has a higher unemployment rate.

This dataset contains 532 rows of information, and a total of 16 columns. Each row represents a year in a specific company, and some key features include company, year, layoffs, and unemployment_rate. Some assumptions that can be made from this dataset is that one company is better than the other because of their employment rates, which could be true, but there are other factors that contribute to that such as wages, work environment and conditions, employment type (salary vs hourly), etc. There are so many other factors that can control how someone decides on what company to work for than what is listed in the dataset. A huge thing that is missing from the dataset could be the wages over time for the employees. We can see that the companies might be billion dollars, but what about what the workers themselves are making vs the CEO?

For data cleaning, initially I wanted to see if there were any duplicate values, and then realized that the features that are present, it wouldn't be bad if there were duplicate values vs having whole duplicate rows. These were not found because looking back at the Kaggle, you can see that there were no duplicates found, but it is always nice to double check.
I decided that the dataset was too large to be working with, especially since I won't be using every feature available, so I decided to create a subset of the original dataset (called dataset), called new_data. This new_data dataset contained only 4 of the 16 features, company, year, layoffs, and revenue_billions_usd. Initially I wanted to work with the revenue and see what kind
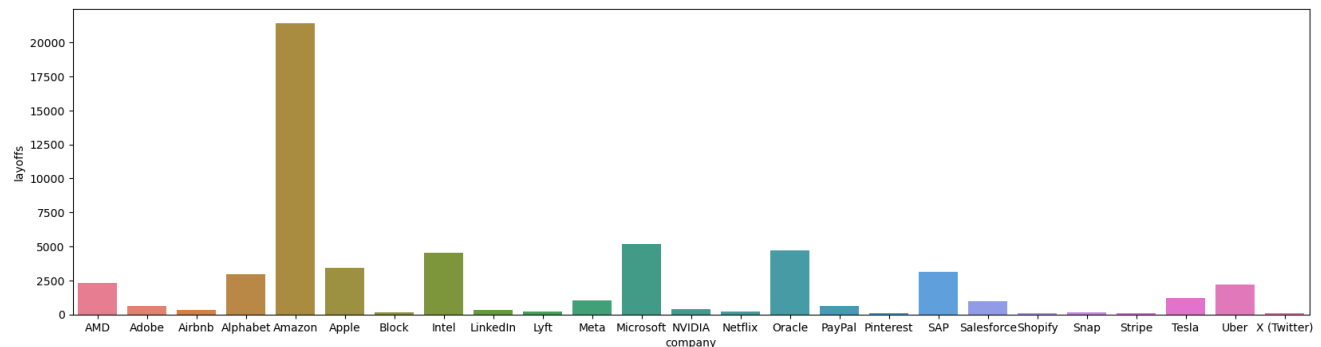
of correlations there are between different features, but I decided against it. So my main features are company, year, and layoffs.
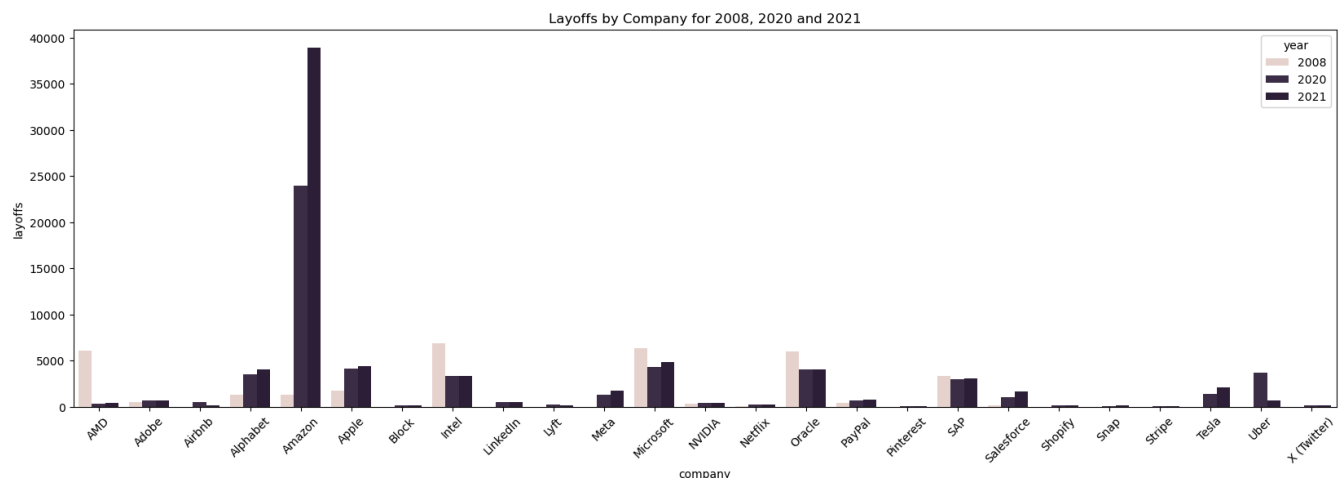
Data Cleaning and Preparation

An assumption that I made was that the dataset was free of duplicates. I did check for null values before doing any other pre-processing steps by scanning through the dataset itself since it is on the relatively smaller side. But it would've been nice to double check and have the computer check for me as well. So from here and creating new_data, I decided to print out the new_data just to double check my work and make sure that there are still 532 observations, but with only 4 features, which there were. I also decided to do another subset but this time having it be with specific years, 2008, 2020 and 2021, and decided to call that 'years'. This will allow me to focus mostly on the data present, and to reduce any confusions on where the data might be wrong. That was it for my preprocessing steps!

Data Understanding and Visualization

For my visualizations, I decided to do a simple bar plot and a more in-depth bar plot. I chose to do a bar plot just to be able to compare the companies and showcase the amount of layoffs that they had overall, and then throughout the years of 2008, 2020, and 2021.

My simple bar plot shows that Amazon (mustard yellow) is the leading company in the amount of layoffs that they accomplished, but it does not specifically show what year that this took place. So now let's take a deeper look into what companies were up to throughout the years 2008, 2020, and 2021.

From the graph above, you can see that Amazon actually has the most layoffs for the year 2021, during the ending season of COVID-19 vs 2008, where they had significantly less layoffs. This shows that in just under 15 years, there were 8 times the amount of layoffs made. From these visuals, we are able to see that companies like Amazon, Alphabet, and Apple had more layoffs during 2021 than they did during the 2008 recession. To conclude that Amazon is the biggest tech company can be misleading, as it only shows a fraction of what goes into deciding that. Especially since there is no data that talks about the amount of technology they have and implemented around the World. It would also be misleading to think that in 2021, that was the bigger crisis because of the amount of layoffs Amazon did solely, but in reality, it was just more drastic than other years. Most of these companies weren't even around or that prominent around 2008, which can also skew the image of how big these companies really are. There are more factors to consider when making judgements!

Limitations, Ethics, and Reflection
Overall, my data does not reflect on the conditions of the company other than their revenue. It touches on the amount of employees it has to begin with and end with, but not the revenue lost, if any. So this doesn't show the economic wellbeing of the company, and the amount of money that it has gained throughout the years. I was more surprised that other companies weren't doing huge amounts of layoffs like Amazon, because I remember Amazon being in the news about it. But honestly, I didn't see other companies being mentioned, just Amazon. That could also create the bias of assuming that Amazon would have the biggest layoff amounts before even looking at the data to see if that's true or not. For further exploration, I would like to see if there is more correlation between features such as the unemployment rate and the revenue amount. If companies aren't making enough money, how will they be able to take on more employees? Do companies with smaller revenue have a higher unemployment rate than a larger company like Amazon?

Thank you for coming to my Ted Talk. 👏

References

- HTML Emojis: https://www.w3schools.com/charsets/ref_emoji_smileys.asp
- Specifically used the "Subsetting Data using Criteria" in order to figure out how to select a certain row and column at the same time. https://datacarpentry.github.io/python-ecology-lesson/03-index-slice-subset.html
- I literally had a brainfart and needed inspiration for my data. I used the Gen AI feature in Google by just typing in my question and Generative AI answered it for me. Hopefully this link works. https://www.google.com/search?q=what+kind+of+visualizations+can+i+do+with+python&rlz=1C1UEAD_enUS1028US1029&oq=what+kind+of+visualizations+can+i+do+with+python&gs_lcrp=EgZjaHJvbWUyBggAEEUYOTIHCAEQIRigATIHCAIQIRigATIHCAMQIRigATIHCAQQIRirAjIHCAUQIRifBdIBCTE5OTg1ajBqN6gCCLACAfEFYdh1MDYL5RvxBWHYdTA2C-Ub&sourceid=chrome&ie=UTF-8
- Trying to remove the error bars in my bar graph. It was annoying me and looked like those fields of cattails. Disgusting. I used Generative AI again because it was the first result that popped up and was relevant. https://www.google.com/search?q=how+to+get+the+lines+off+of+the+barplot+in+python+in+seaborn&sca_esv=b74357875f8bb801&rlz=1C1UEAD_enUS1028US1029&biw=1920&bih=945&sxsrf=ANbL-n4iswVYfjtJVyrZ9cF2OZevy3rVSQ%3A1772063179954&ei=y4mfafv1OdPFkPIP1Nq3mQQ&ved=0ahUKEwi7h_Hi6fWSAxXTIkQIHVTtLUMQ4dUDCBE&uact=5&oq=how+to+get+the+lines+off+of+the+barplot+in+python+in+seaborn&gs_lp=Egxnd3Mtd2l6LXNlcnAiPGhvdyB0byBnZXQgdGhlIGxpbmVzIG9mZiBvZiB0aGUgYmFycGxvdCBpbiBweXRob24gaW4gc2VhYm9ybjIHECEYoAEYCjIHECEYoAEYCjIHECEYoAEYCjIHECEYoAEYCkjuSFCVCFimR3ADeACQAQKYAb8CoAG-SaoBCDAuNTguMi4xxuAEDyAEA-AEBmAI9oALXSMICChAAGLADGNYEGEfCAgsQABiABBiRAhiKBcICChAAGIAEGEMYigXCAg4QLhiABBixAxjRAxjHAcICCxAuGIAEGNEDGMcBwgIOEAAYgAQYsQMYgwEYigXCAg4QLhiABBixAxiDARiKBcICBRAuGIAEwgIIEC4YgAQYsQPCAgUQABiABMICCBAAGIAEGLEDwgINEAAYgAQYsQMYQxiKBcICBhAAGBYYHsICCxAAGIAEGlYDGIoFwgIFEAAY7wXCAggQABiABBiiBMICBAAGKIEGIkFwgIFECEYoAHCAgUQIRirAsICBRAhGJ8FmAMAiAYBkAYIkgcGMy41NC40oAeXwgOyBwYwLjU0LjS4B8hIwgclMC43LjQ2LjjIB84CgAgA&sclient=gws-wiz-serp
- Was trying to figure out the parameters for the seaborn bar plot. I don't think I even ended up using this. https://seaborn.pydata.org/generated/seaborn.barplot.html
- I also did use Copilot in VS Code, it kept popping up but it is referenced in my code on the sections I did use it on! Mostly just for my nice bar graph.