

LAYOFF PROJECT



Contents

1. Introduction
2. Problem Statement
3. Dataset and its Attribute
4. Data Cleaning and Preprocessing
5. EDA
6. Conclusion

Introduction

In the contemporary corporate landscape, layoffs have become significant, influenced by various economic, industrial, and company-specific factors. Our project explored a comprehensive layoffs dataset, aiming to uncover patterns and insights through rigorous Exploratory Data Analysis (EDA). Leveraging a diverse set of analytical tools and libraries, such as Pandas for data manipulation, Seaborn and Matplotlib for data visualization, and Plotly for interactive plots, our analysis delved into the multifaceted aspects of layoffs. We employed statistical methods, including linear regression from SciPy, to interpret the data meaningfully. The utilization of advanced techniques like K-Means clustering and standard scaling further enriched our understanding of the dataset's intricacies.

This report aims to present our findings, offering a detailed view of the layoff landscape across various industries and geographies. The insights drawn from this analysis are intended to provide valuable inputs for stakeholders, including business leaders, policymakers, and economic analysts, offering a data-driven perspective on the patterns and trends in layoffs worldwide. This introduction sets the stage for your detailed EDA findings, highlighting the tools used and the scope of your analysis.

Problem Statement

The primary aim is to analyze and understand the patterns and trends in layoffs across various industries and locations. The focus is on identifying key factors contributing to layoffs and how different sectors and regions are affected. The analysis aims to provide a comprehensive understanding of the layoff landscape, identifying industries and areas most affected, trends over time, and the potential impact of financial factors like funds raised. In this Project, We have a dataset consisting of 2103 entries for total layoffs, 2044 for percentage laid off, and 2820 for funds raised, indicating the presence of missing values in these fields.

Dataset and Attributes

Dataset link: <https://www.kaggle.com/datasets/swaptr/layoffs-2022>

The dataset is based on the EDA file, is centered around layoffs from various companies. It contains a range of attributes that offer insights into these layoffs' scale, industry distribution, and geographical spread. Here's an extensive explanation of the dataset and its attributes:

- **Company:** This column lists the companies that have reported layoffs. It directly references the organizations affected and is crucial for identifying industry trends and company-specific issues.
- **Location:** This attribute specifies the geographical location of the companies, typically given as city names. The location data is essential for understanding the geographic distribution of layoffs and identifying if certain regions are more affected than others.
- **Industry:** The industry column categorizes each company into a specific sector, such as Crypto, Real Estate, Data, etc. This classification is vital for analyzing which industries are experiencing higher rates of layoffs and understanding sector-specific trends.
- **Total Laid Off:** This numeric field represents the total number of employees laid off by the respective companies. This is a critical metric in the dataset as it quantifies the impact of layoffs at a company level.

- **Percentage Laid Off:** Alongside the total number, this column provides the percentage of the laid-off workforce. This attribute offers a relative sense of the layoff's impact, which is especially useful for comparing companies of different sizes.
- **Date:** The date of the layoffs or the report gives a temporal context to the data. It allows for the analysis of layoffs over time and can help identify any temporal patterns or trends.
- **Stage:** This column likely refers to the stage of the company, such as startup, growth, maturity, etc., though the specific categories need to be detailed in the provided excerpts. The stage of a company can be an essential factor in understanding its stability and the reasons behind the layoffs.
- **Country:** This attribute specifies the country in which the company operates. Like the location, it's crucial for analyzing the geographical spread and impact of layoffs on a global scale.
- **Funds Raised:** This numeric field indicates the amount of funds raised by the company, presumably over its lifetime. This financial metric can be critical in understanding the financial health of the company and its potential correlation with layoffs.

Data cleaning and preprocessing

The dataset you provided underwent several data cleaning and preprocessing steps to prepare it for analysis. Here is a detailed explanation of these methods:

	company	location	industry	total_laid_off	percentage_laid_off	Date	stage	country	funds_raised
0	OpenSea	New York City	Crypto	NaN	0.50	03-11-2023	Series C	United States	427.0
1	Viasat	San Diego	Other	800.0	0.10	02-11-2023	Post-IPO	United States	366.0
2	OpenSpace	SF Bay Area	Construction	NaN	NaN	02-11-2023	Series D	United States	200.0
3	Orchard	New York City	Real Estate	NaN	NaN	02-11-2023	Series D	United States	472.0
4	Informatica	SF Bay Area	Data	545.0	0.10	01-11-2023	Post-IPO	United States	NaN
...
3143	Service	Los Angeles	Travel	NaN	1.00	16-03-2020	Seed	United States	5.1
3144	HopSkipDrive	Los Angeles	Transportation	8.0	0.10	13-03-2020	Unknown	United States	45.0
3145	Panda Squad	SF Bay Area	Consumer	6.0	0.75	13-03-2020	Seed	United States	1.0
3146	Tamara Mellon	Los Angeles	Retail	20.0	0.40	12-03-2020	Series C	United States	90.0
3147	EasyPost	Salt Lake City	Logistics	75.0	NaN	11-03-2020	Series A	United States	12.0

3148 rows x 9 columns

- **Handling Missing Values:**

0.1 Initial Assessment

- To check for missing values in the dataset, we used the `df.isna().sum()` method.
- The method revealed missing values in several columns, including:
 - 'total_laid_off'
 - 'percentage_laid_off'
 - 'funds_raised'

0.2 Dropping Rows with Missing Values

- Rows with missing values in crucial columns were dropped from the dataset.
 - * **Columns:** 'total_laid_off', 'location', 'industry', 'stage'
 - * **Method:** `df.dropna(subset=['total_laid_off', 'location', 'industry', 'stage'], inplace=True)`

0.3 Filling Missing Values

- For columns 'percentage_laid_off' and 'funds_raised', missing values were filled with the median of their respective columns.
 - * **Columns:** 'percentage_laid_off', 'funds_raised'
 - * **Method:** `df['column'].fillna(median_value, inplace=True)`

	company	location	industry	total_laid_off	percentage_laid_off	Date	stage	country	funds_raised
1	Viasat	San Diego	Other	800.0	0.10	02-11-2023	Post-IPO	United States	366.0
4	Informatica	SF Bay Area	Data	545.0	0.10	01-11-2023	Post-IPO	United States	NaN
5	Splunk	SF Bay Area	Data	500.0	0.07	01-11-2023	Post-IPO	United States	2400.0
6	Faire	SF Bay Area	Retail	250.0	0.20	01-11-2023	Series G	United States	1700.0
7	Sana Benefits	Austin	HR	73.0	0.50	01-11-2023	Series B	United States	106.0
...
3142	Help.com	Austin	Support	16.0	1.00	16-03-2020	Seed	United States	6.0
3144	HopSkipDrive	Los Angeles	Transportation	8.0	0.10	13-03-2020	Unknown	United States	45.0
3145	Panda Squad	SF Bay Area	Consumer	6.0	0.75	13-03-2020	Seed	United States	1.0
3146	Tamara Mellon	Los Angeles	Retail	20.0	0.40	12-03-2020	Series C	United States	90.0
3147	EasyPost	Salt Lake City	Logistics	75.0	NaN	11-03-2020	Series A	United States	12.0

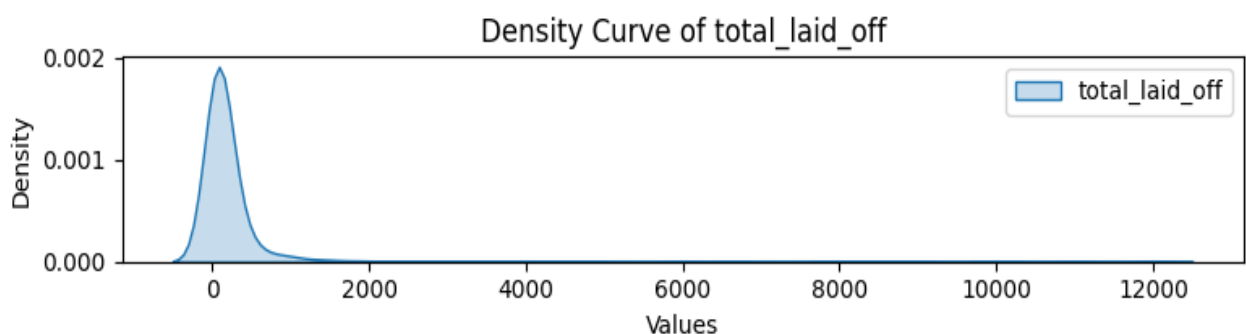
2099 rows x 9 columns

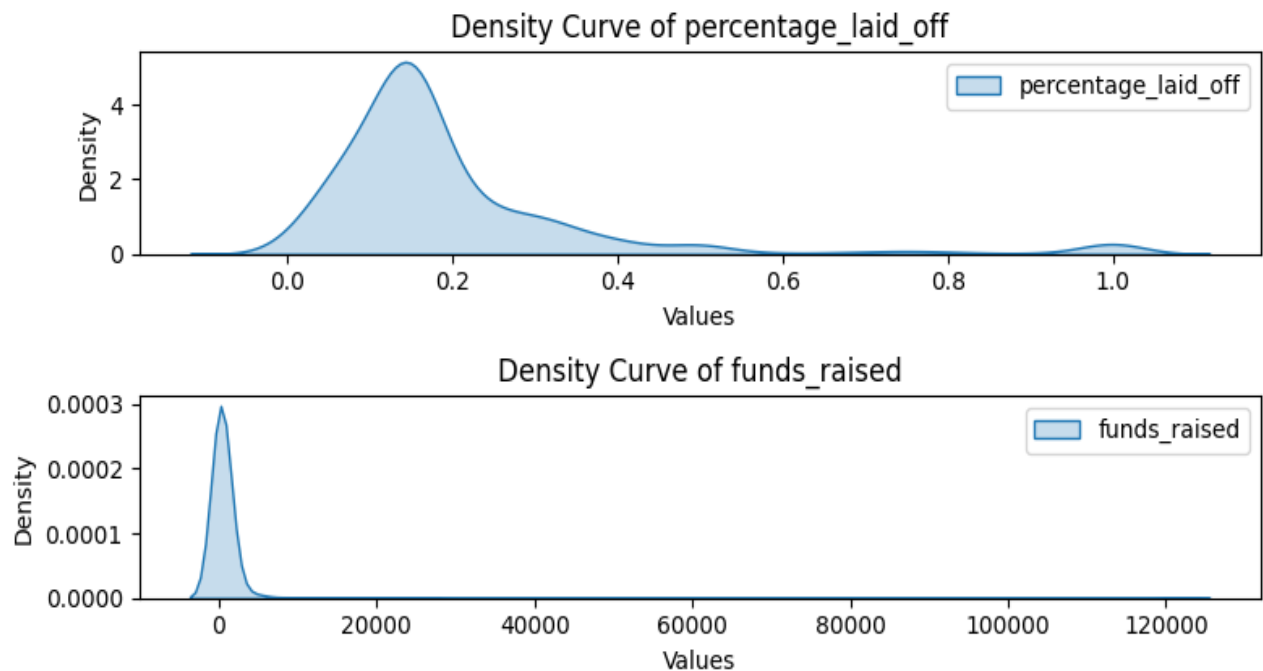
0.4 Data Transformation

- **Conversion to Datetime:** The 'Date' column was converted to a datetime format using `pd.to_datetime(df['Date'])`. This is important for any time series analysis, as it ensures that dates are correctly formatted and recognized.
- **Indexing by Date:** The dataset was then indexed by the 'Date' column to facilitate time-based analysis. This step is crucial for chronological analysis and time series forecasting.

Exploratory Data Analysis (EDA)

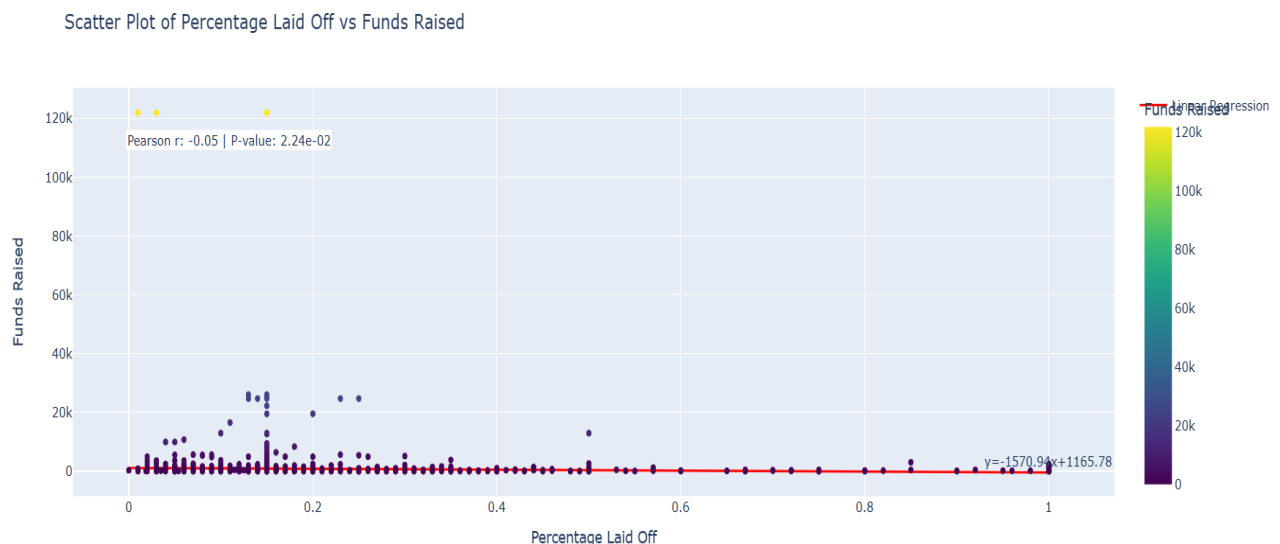
- **Density Curves:**





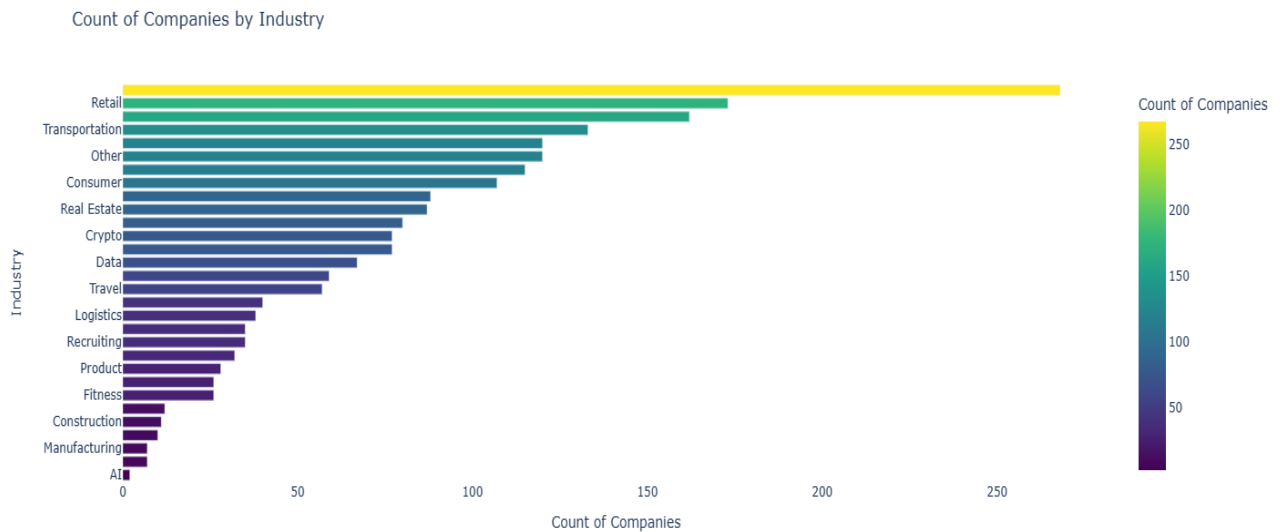
- The graphs display density curves for total layoffs, the percentage of employees laid off, and funds raised. The first graph shows that most companies have a low number of layoffs, with a sharp peak at the lower end. The second graph indicates a similar skew towards lower percentages of layoffs. The third graph suggests most companies have raised a relatively small amount of funds, again with a peak at the lower end, indicating that higher fund-raising is less common.

• Linear regression



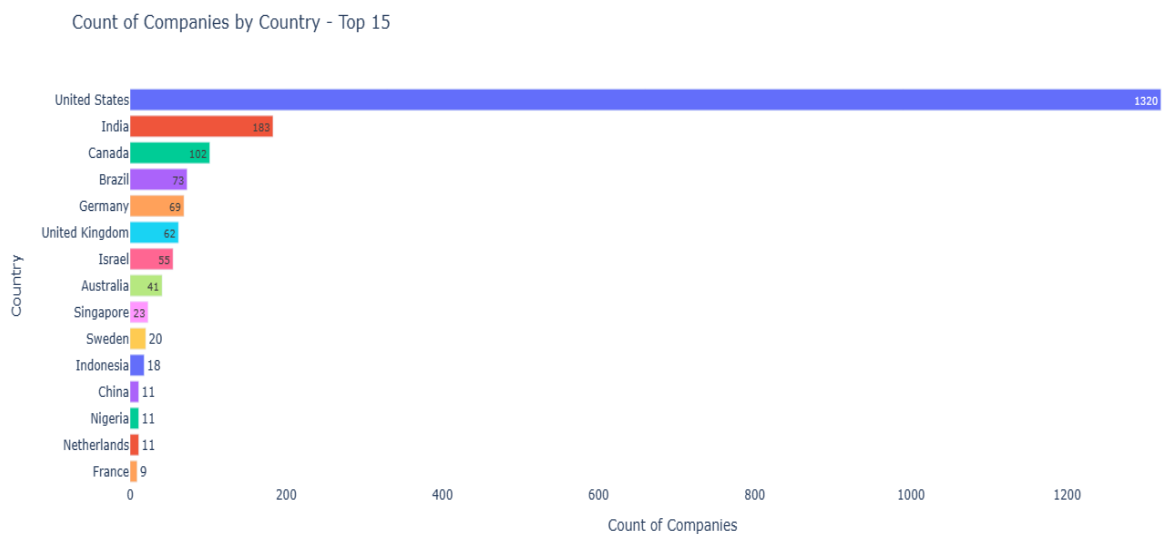
- Based on the scatter plot and the accompanying statistical measures, there is a weak negative relationship between the percentage of the workforce laid off and the funds raised by companies. The p-value is 2.24e-02 (or 0.0224), which suggests that the relationship observed is statistically significant at the 5% significance level (since the p-value is less than 0.05). However, the correlation is very weak. While the relationship is statistically significant, the correlation coefficient is so close to zero that it suggests no practical linear relationship. This indicates that within this dataset, the amount of funds a company raises is not a strong predictor of the percentage of its workforce that may be laid off.

- Count of companies by industry



- The graph depicts the number of companies by industry, with the retail and transportation industries having the highest counts. In contrast, AI and Manufacturing have the lowest, indicating the industry-wise distribution of companies in the dataset.

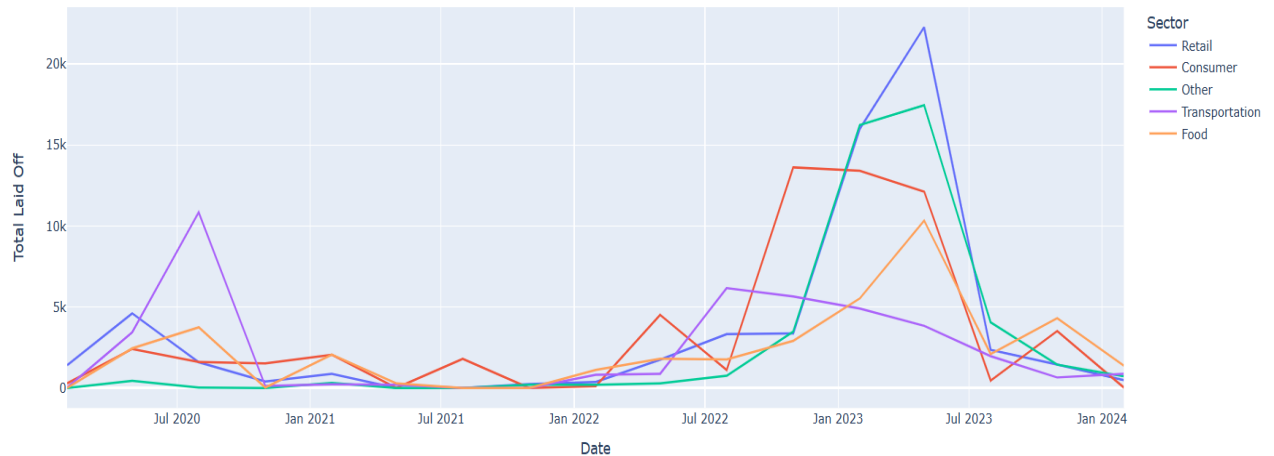
- Count of companies by countries



- The bar chart shows the number of companies by country and displays the top 15 countries. The United States leads substantially, followed by India and Canada. This illustrates the dominance of the U.S. in this dataset and highlights the global distribution of companies, with representation across North America, Europe, Asia, and South America, indicating the diverse multinational presence of companies considered in the dataset.

- Layoff Trends in top 5 sectors

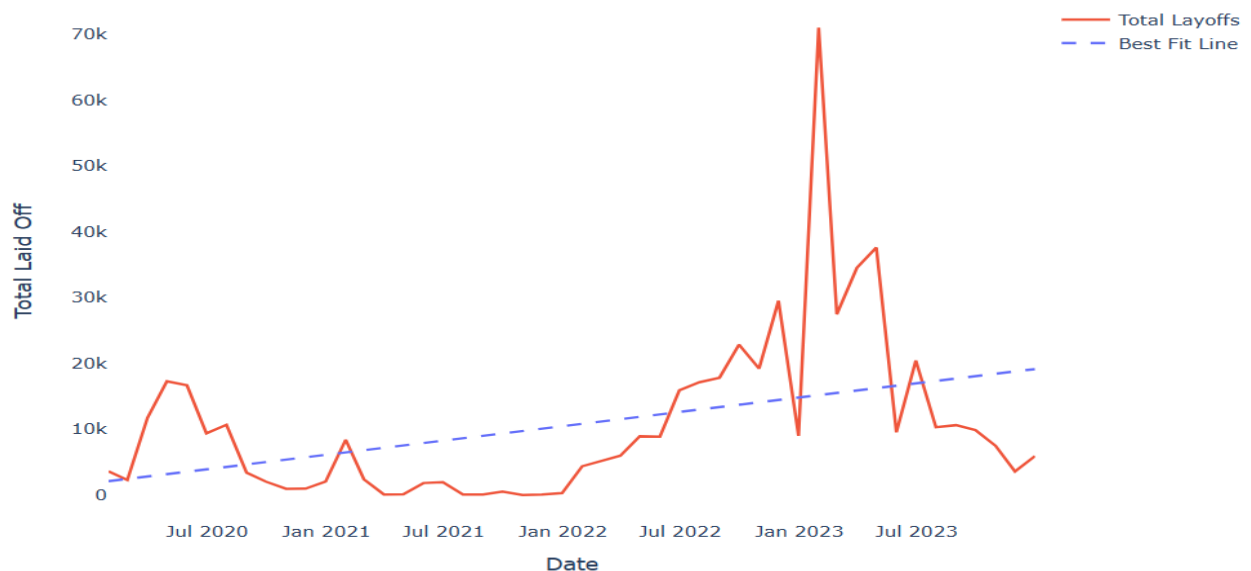
Layoff Trends in Top 5 Sectors Over Time (Reduced to 3 Months Intervals)



- The line graph illustrates layoff trends over time in three-month intervals for the top five sectors. It shows significant layoff fluctuations, with peaks and troughs indicating varying levels of layoffs in different quarters. Transportation and Retail sectors exhibit the most pronounced changes, suggesting seasonal or cyclical impacts, as well as potential market-driven events affecting these industries' workforce numbers

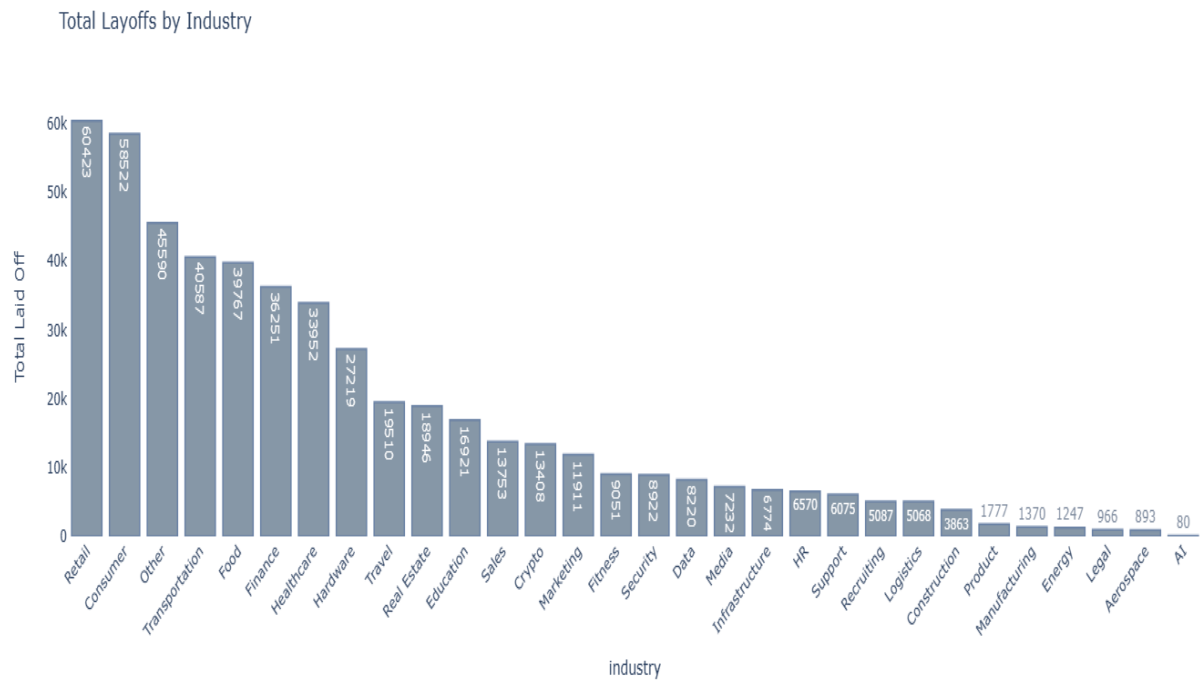
• Monthly Layoffs

Monthly Total Layoffs



- The graph represents monthly total layoffs with a significant spike, indicating a peak in layoffs at a particular point in time. The blue dashed line represents the best fit line, showing the overall trend excluding the outlier. Despite the peak, the trend line suggests a relatively stable pattern of layoffs over the observed period, indicating that the spike may be an anomaly rather than a consistent trend

• Total layoffs by industry



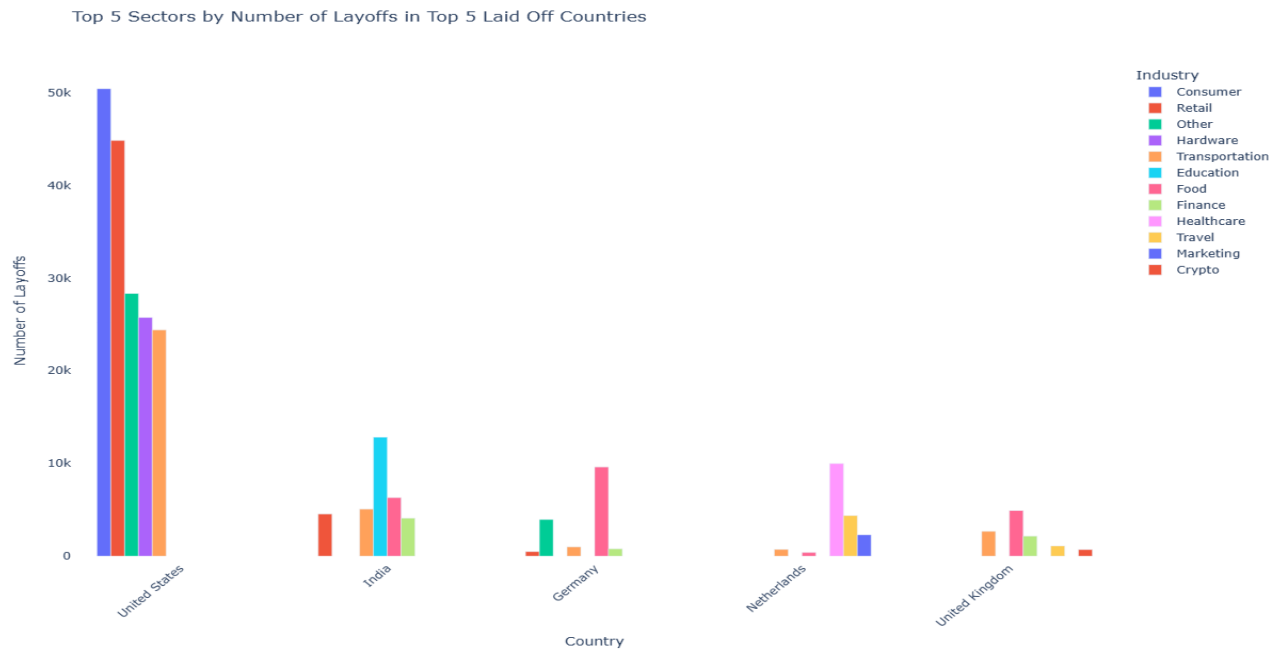
- The bar graph displays total layoffs by industry, revealing a descending order of impacted sectors. Retail leads with the highest layoffs, followed closely by Consumer and Other categories. The trend shows a steep decline in layoffs as industries transition from Transport to AI, suggesting that some sectors are more vulnerable to layoffs than others, possibly due to economic, technological, or sector-specific challenges.

• Pivot Table

industry	AI	Aerospace	Construction	Consumer	Crypto	Data	Education	Energy	Finance	Fitness	Food	HR	Hardware	Healthcare	Infrastructure	Legal	Logistics	Manufacturing
country																		
Argentina	0	0	0	0	180	0	0	0	53	0	0	0	0	0	0	0	0	0
Australia	0	0	0	110	264	0	0	88	543	22	890	186	0	18	80	0	27	0
Austria	0	0	0	0	270	0	300	0	0	70	0	0	0	0	0	0	0	0
Brazil	0	0	0	381	190	0	319	0	3671	467	675	68	0	311	0	0	800	0
Bulgaria	0	0	0	0	0	0	0	0	0	0	0	0	0	0	120	0	0	0
Canada	0	0	293	22	102	45	612	47	777	0	913	9	50	146	30	30	65	0
Cayman Islands	0	0	0	0	1000	0	0	0	0	0	0	0	0	0	0	0	0	0
Chile	0	0	0	0	0	0	0	0	0	0	250	0	0	75	0	0	0	0
China	0	0	0	2445	575	0	0	0	0	0	0	0	900	500	0	0	0	0
Colombia	0	0	100	0	0	0	0	0	0	0	0	30	0	0	0	0	0	0
Denmark	0	0	0	15	0	0	40	0	0	0	150	0	0	0	0	0	0	0
Estonia	0	0	0	0	0	0	0	0	24	0	0	0	0	0	0	0	0	0
Finland	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
France	0	0	0	0	0	0	0	0	0	0	17	200	0	0	0	0	0	0
Germany	0	0	0	0	0	259	70	0	791	155	9618	68	0	81	0	0	0	0
Ghana	0	0	0	0	0	0	0	0	0	0	0	0	0	150	0	0	0	0
Hong Kong	0	0	0	0	30	0	0	0	400	0	0	0	0	0	0	0	0	0
India	0	0	140	3558	284	217	12829	0	4097	1100	6315	80	0	1747	0	0	680	0

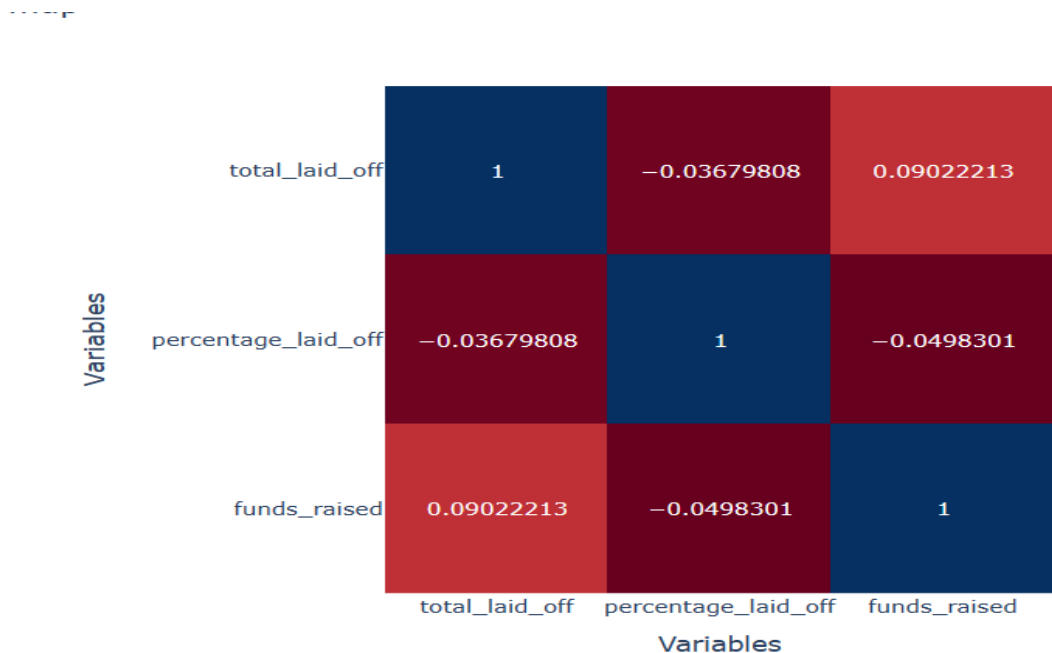
- It aggregates data by summing up layoffs, with missing values filled with zero. The table is then styled to format the numbers without decimal places. This pivot table allows for a clear, concise comparison of layoffs across different industries and countries, highlighting areas with higher and lower job losses.

- **Top 5 sectors with top 5 countries**



- The bar chart illustrates layoffs within the top five sectors across the top five countries with the highest layoffs. The United States has the most significant number of layoffs across almost all sectors, particularly in the Consumer sector. The chart demonstrates the variation and distribution of layoffs by sector within each country, with other countries showing notably fewer layoffs compared to the U.S. across corresponding sectors.

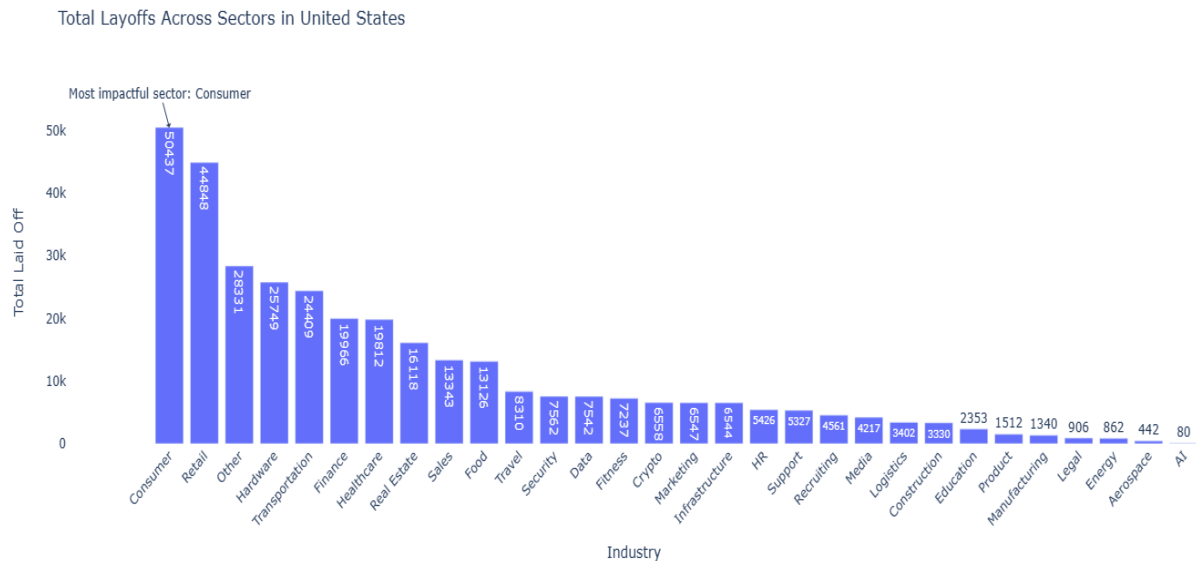
- **Correlation Matrix Heatmap**



- The heatmap represents a correlation matrix, visualizing the relationship between total layoffs, percentage laid off, and funds raised. The color intensity indicates the strength and direction of the correlation. Dark blue shows a positive correlation, while dark red indicates a negative correlation. The matrix shows very weak correlations between the variables,

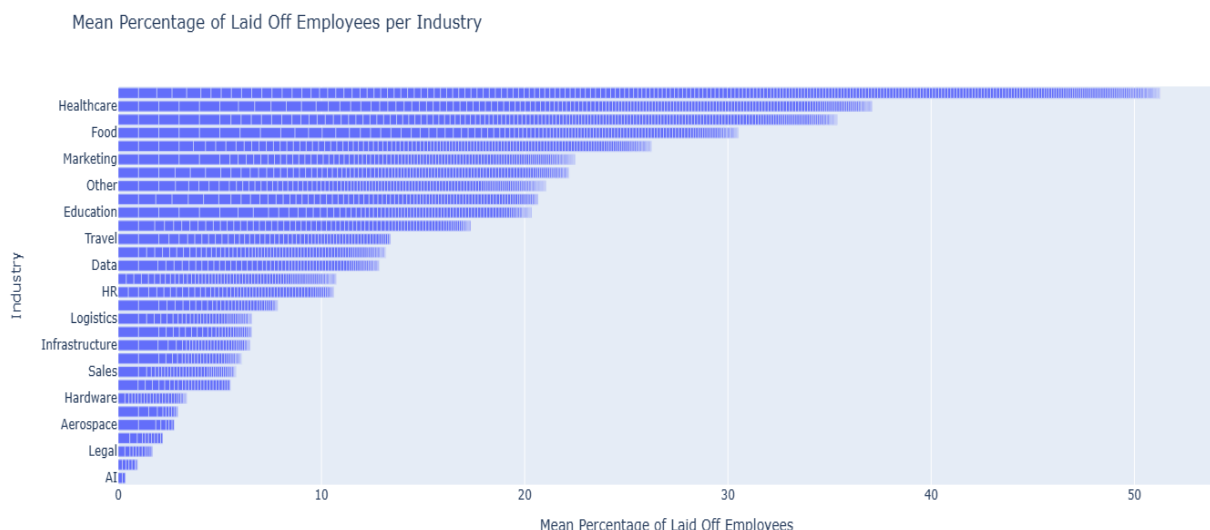
suggesting that within this dataset, these variables do not have strong linear relationships with each other.

• Total layoff across USA



- The bar chart displays total layoffs in various sectors within the United States, with the Consumer sector experiencing the highest number of layoffs, followed by Retail and Other. This visual representation emphasizes the disproportionate impact on the Consumer sector compared to others, like AI and Aerospace, which had the least number of layoffs. This data could suggest sector-specific economic challenges or transformations impacting employment.

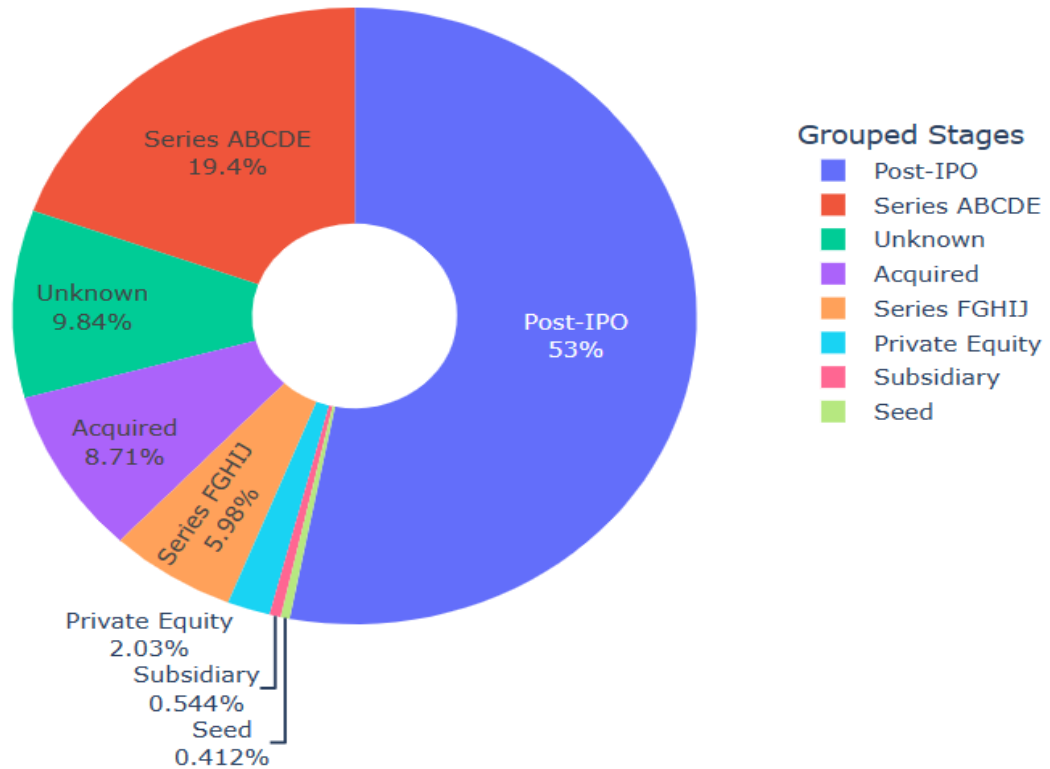
• Mean percentage of layoffs:



- The bar chart visualizes the mean percentage of employees laid off across various industries. Healthcare and Food industries have the highest mean percentage of layoffs, suggesting a significant workforce reduction in these sectors. The bars represent the average layoffs per industry, providing a quick comparison of how different sectors have been affected. Shorter bars for industries like Legal and AI indicate a lower mean percentage of layoffs in these fields.

- Pie chart

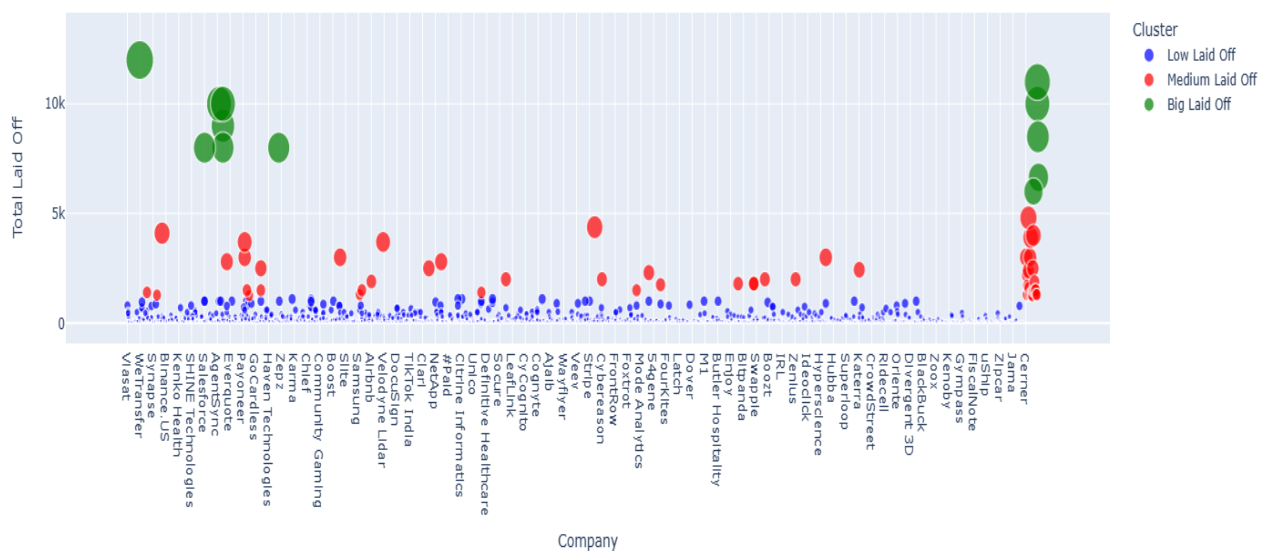
Distribution of Total Layoffs by Grouped Stages



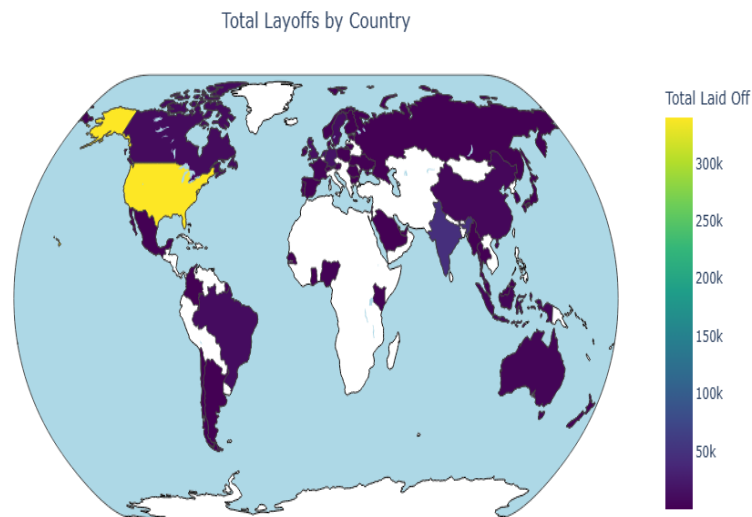
- The pie chart illustrates the distribution of total layoffs by company stages, with a dominant share of layoffs occurring in post-IPO companies, accounting for 53

- Clustering

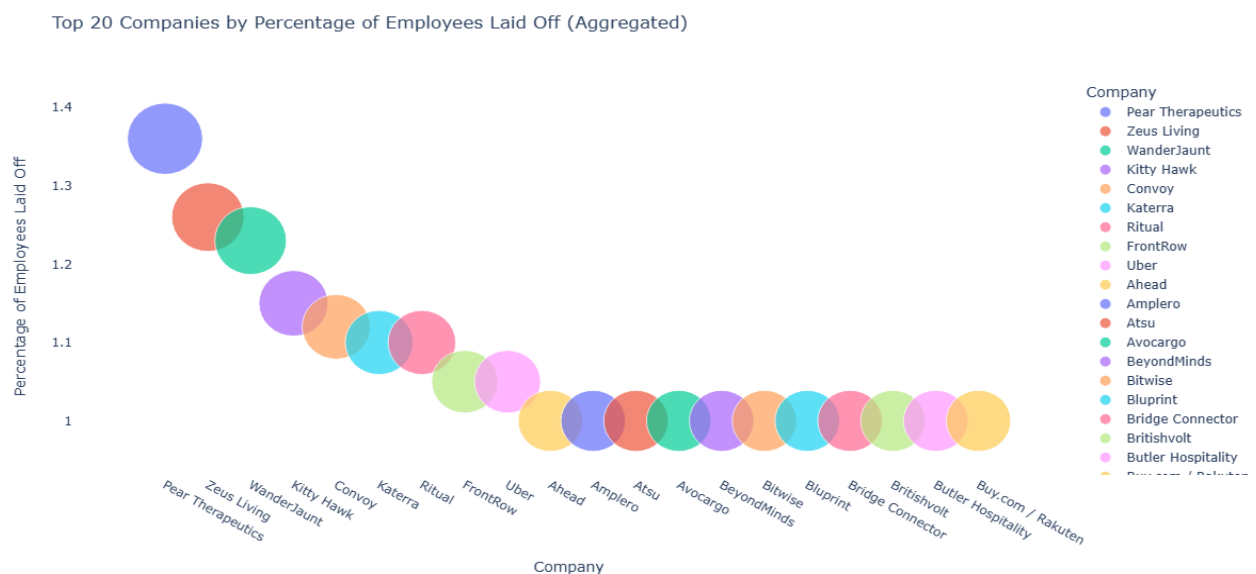
Company Clusters by Total Laid Off (Bubble Chart)



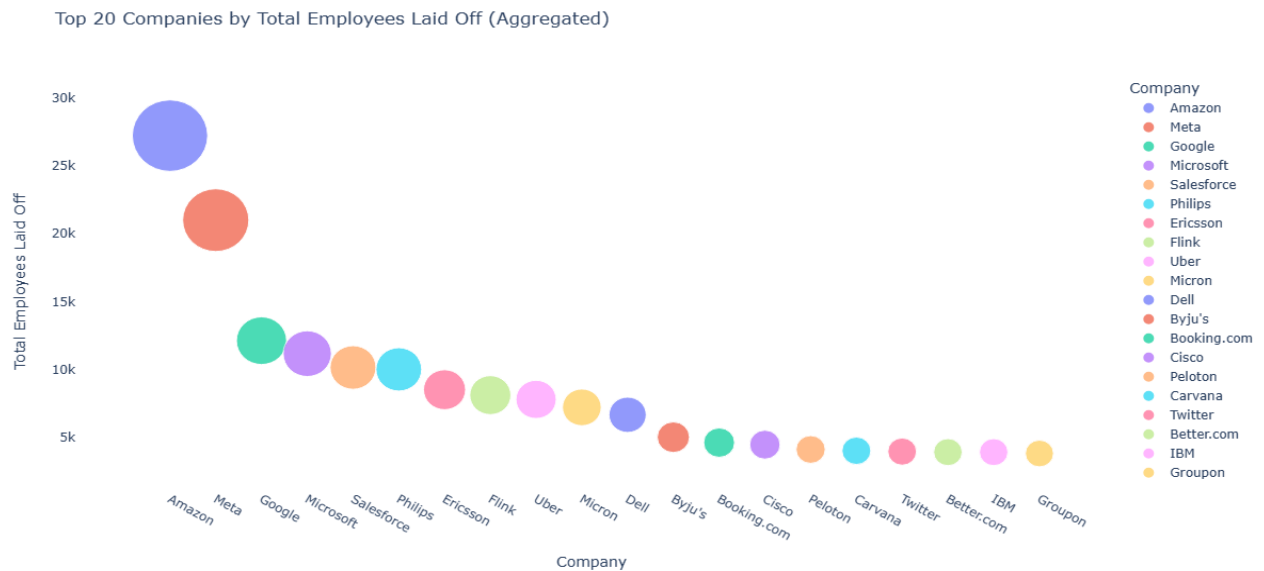
- The bubble chart categorizes companies into clusters based on the total number of layoffs, with the size of each bubble reflecting the layoff magnitude. Blue represents companies with low layoffs, red for medium, and green indicates companies with significant layoffs. The chart highlights that while most companies have a low to medium number of layoffs, a few companies, particularly those represented by larger green bubbles, have experienced a disproportionately high number of layoffs.
- Worldmap:**



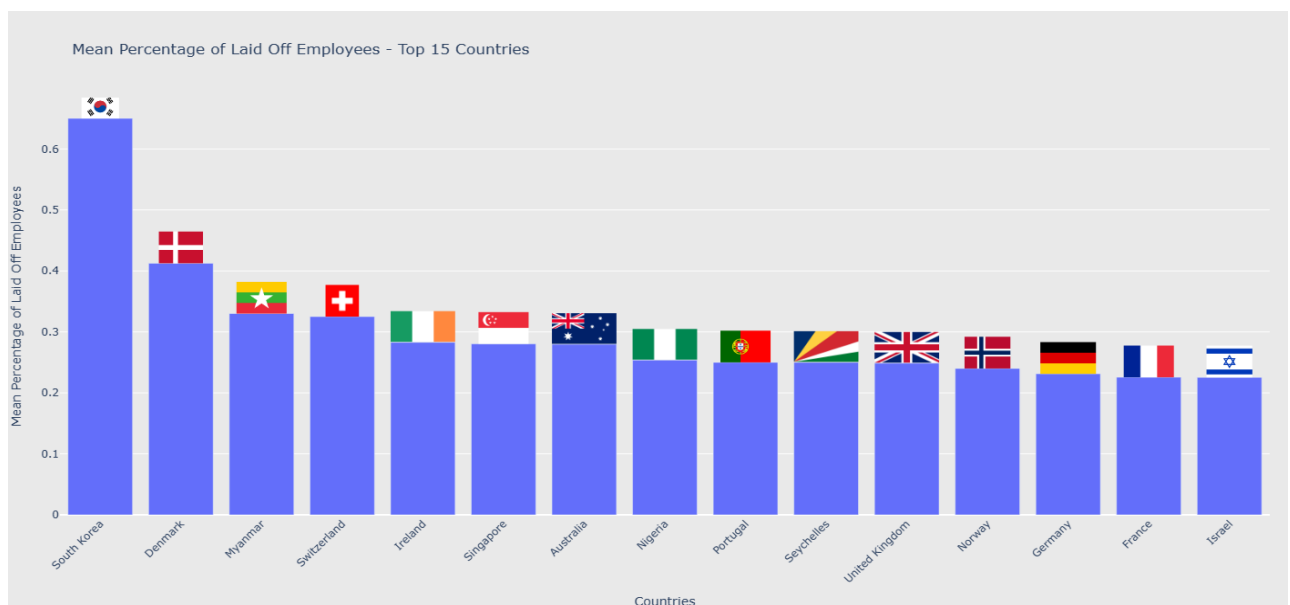
- The world map visualizes total layoffs by country, with color intensity corresponding to the number of layoffs: darker colors indicate higher layoffs. The United States stands out with the highest layoffs, followed by various shades across other countries, reflecting lower numbers. This map highlights the global impact of layoffs, showing that while some countries experience high unemployment rates, others are less affected, providing a geographical perspective on the economic challenges faced worldwide.
- Top 20 companies by percentage laid-off:**



- The bubble chart represents the top 20 companies by the percentage of employees laid off, with the bubble size indicating the layoff rate. Larger bubbles denote a higher percentage of workforce reductions. The chart is color-coded to differentiate between companies and showcases a range from the highest percentage at the top to lower percentages as the bubbles decrease in size. This visualization highlights the companies most affected by layoffs in a comparative context.
- **Top 20 companies by total employees laid off:**



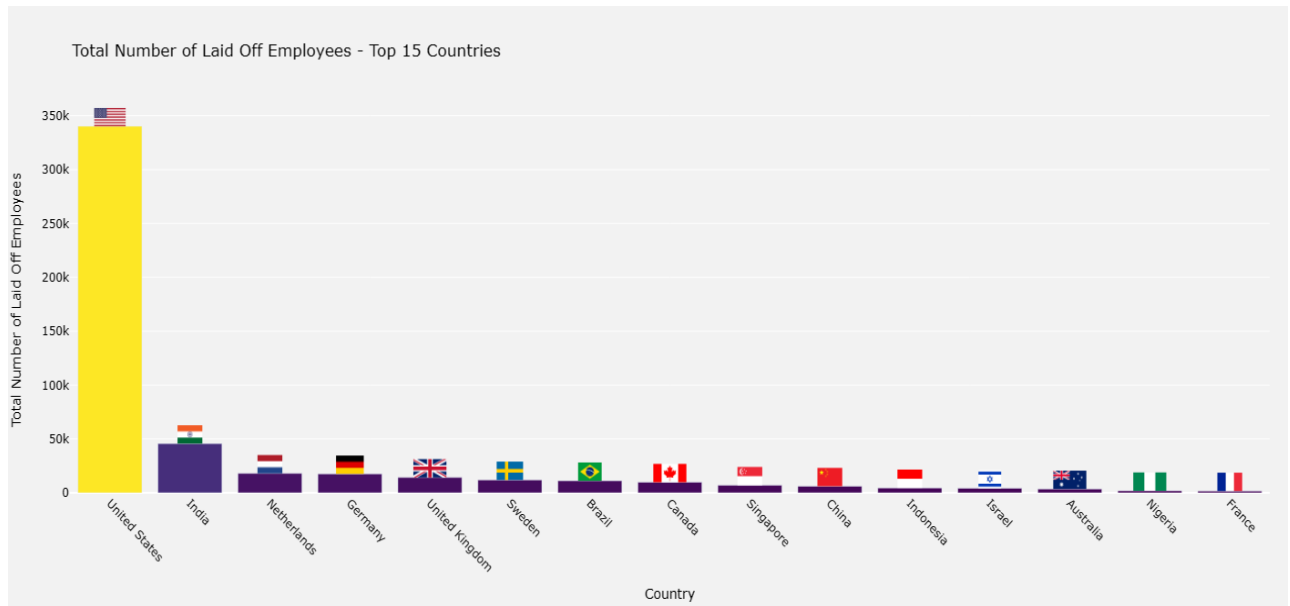
- The bubble chart displays the top 20 companies by total number of employees laid off. The size of each bubble corresponds to the total layoffs, with Amazon and Meta having the largest bubbles, indicating the highest number of layoffs. The chart offers a visual comparison among leading companies, showing a significant variation in the scale of layoffs, which decreases markedly among companies with smaller bubbles, like IBM and Groupon, towards the end of the chart.
- **Mean percentage of total laid off:**



- The bar chart illustrates the mean percentage of employees laid off in the top 15 countries. South Korea shows the highest mean layoff rate, significantly more than the other countries

listed. The chart displays a descending order of mean layoff percentages, with countries like Israel, France, and Germany having the lowest rates among the top 15.

- **Total laid off- Top 15 countries**



- The bar chart shows the total number of employees laid off in the top 15 countries, with the United States far exceeding other countries in layoffs. India, the Netherlands, and Germany follow, but with significantly lower numbers. The bars are color-coded with the respective national flags, emphasizing the stark contrast between the United States and other countries. The data suggests a considerable geographical disparity in the impact of layoffs on the workforce.

- **Conclusion:**

- The exploratory data analysis (EDA) on layoffs revealed significant insights. It identified a substantial difference between the means of funds raised and the percentage of employees laid off, suggesting a complex relationship between a company's financial influx and its layoff decisions.
- The recent increase in layoff trends is because of multiple factors involving the COVID-19 pandemic, Russia, the Ukraine war, and the collapse of a few financial institutions. The instability of supply and demand in the markets, affecting the company's revenue, led to budget cuts to maximize profits through layoffs. Through this extensive analysis, we will be well-informed in making career or higher education choices wisely. The business point of view is where we can have an idea of our wise investments in the sectors or the company.
- The analysis also highlighted industry-specific trends, where certain sectors experienced higher layoffs. This study was crucial in understanding the dynamics of layoffs in the corporate world, providing a data-driven perspective on how economic and industry-specific factors influence workforce reductions. This understanding aids in better preparing for and mitigating the impacts of such challenging corporations.