Exploratory Data Analysis (EDA) Project Documentation

Project Title: Exploratory Data Analysis on Cleaned Layoffs Dataset

Project Overview:

This project aims to conduct an exploratory data analysis (EDA) on a cleaned dataset related to layoffs across various companies. The dataset contains critical information about layoffs such as company names, industry, total laid-off employees, percentage laid off, dates, stages, and funds raised by companies. The EDA focuses on discovering insights and identifying trends in company layoffs, industries, countries, and time periods, using SQL queries to derive meaningful insights.

Dataset Overview:

The dataset, which has been cleaned in the previous step, consists of several important columns:

- company: Name of the company
- total_laid_off: Total number of employees laid off
- percentage_laid_off: Percentage of the company workforce laid off
- date: Date of the layoffs
- industry: Industry to which the company belongs
- stage: Stage of the company (e.g., Series A, Post-IPO)
- country: Country where the company is based
- funds_raised_millions: Funds raised by the company in millions

EDA Steps and SQL Queries:

1. Initial Check of Cleaned Dataset

The first step is to verify that the cleaned data is ready for analysis.

```
SELECT *
FROM layoffs_staging2;
```

This query displays the entire dataset to ensure that all fields have been cleaned and there are no issues like missing or corrupted data.

2. Maximum Layoffs and Percentage

To understand the largest layoffs, we begin by checking the maximum number of employees laid off and the maximum percentage of layoffs.

```
SELECT MAX(total_laid_off), MAX(percentage_laid_off)
FROM layoffs_staging2;
```

- **Insight**: This query reveals that one company laid off over 12,000 employees, and some companies had a 100% layoff, meaning they completely shut down.

3. Companies with 100% Layoff

Next, we identify the companies that had 100% layoffs (i.e., the company completely closed down).

```
SELECT *
FROM layoffs_staging2
WHERE percentage_laid_off = 1;
```

- **Insight**: This query shows companies that had to lay off their entire workforce.

4. Highest Layoff Companies with 100% Layoff

To find out which company laid off the highest number of employees while completely shutting down, we sort the result by the total number of layoffs.

```
SELECT *
FROM layoffs_staging2
WHERE percentage_laid_off = 1
ORDER BY total_laid_off DESC;
```

- **Insight**: This reveals that Katerra, a construction company, had the highest layoffs with a 100% layoff rate.

5. Companies with Highest Funds Raised and 100% Layoff

We investigate which company raised the highest amount of funds but still shut down.

```
SELECT *
FROM layoffs_staging2
WHERE percentage_laid_off = 1
ORDER BY funds_raised_millions DESC;
```

- **Insight**: The company 'Britishvolt' raised over 2.4 billion dollars in funding but still went bankrupt.

6. Total Layoffs by Company

We calculate the total layoffs for each company to understand which companies had the most layoffs overall.

```
SELECT company, SUM(total_laid_off)
FROM layoffs_staging2
GROUP BY company
ORDER BY 2 DESC;
```

- **Insight**: Companies like Amazon, Meta, and Uber show up at the top with thousands of layoffs.

7. Date Range of Layoffs

To examine the date range of layoffs, we check the minimum and maximum dates from the dataset.

```
SELECT MIN(`date`), MAX(`date`)
FROM layoffs_staging2;
```

- **Insight**: Layoffs in the dataset span from 2020 to 2023, reflecting trends over this period.

8. Industries with the Most Layoffs

To understand which industries were hardest hit, we sum the layoffs by industry.

```
SELECT industry, SUM(total_laid_off)
FROM layoffs_staging2
GROUP BY industry
ORDER BY 2 DESC;
```

- **Insight**: Industries like Consumer, Retail, and Transportation were hit the hardest, with thousands of employees laid off.

9. Layoffs by Country

Next, we explore which countries had the highest number of layoffs.

```
SELECT country, SUM(total_laid_off)
FROM layoffs_staging2
GROUP BY country
ORDER BY 2 DESC;
```

- **Insight**: The United States had the most layoffs, followed by countries like India, Netherlands, and Sweden.

10. Layoffs by Year

We analyze layoffs by year to see how they changed over time.

```
SELECT YEAR(`date`), SUM(total_laid_off)
FROM layoffs_staging2
GROUP BY YEAR(`date`)
ORDER BY 1 DESC;
```

- **Insight**: The year 2022 saw the highest number of layoffs, followed by 2023, highlighting the impact of global economic downturns.

11. Stage-wise Layoffs

To see the layoffs at different company stages (such as Series A, Post-IPO), we group by stage and total layoffs.

```
SELECT stage, SUM(total_laid_off)
FROM layoffs_staging2
GROUP BY stage
ORDER BY 2 DESC;
```

- **Insight**: Post-IPO companies like Amazon and Google had the highest layoffs, followed by companies in the Series C and D stages.

12. Rolling Sum of Layoffs (Month-wise)

We calculate the rolling sum of layoffs over months using a window function.

```
WITH rolling_total AS (
   SELECT SUBSTRING(`date`, 1, 7) AS `MONTH`, SUM(total_laid_off)
AS total_off
   FROM layoffs_staging2
   WHERE SUBSTRING(`date`, 1, 7) IS NOT NULL
   GROUP BY SUBSTRING(`date`, 1, 7)
   ORDER BY 1 ASC
)
SELECT `MONTH`, total_off, SUM(total_off) OVER(ORDER BY
`MONTH`) AS rolling_total
FROM rolling_total;
```

- **Insight**: This shows the cumulative sum of layoffs month by month, helping us see the overall trend in layoffs over time.

13. Top 5 Companies with Highest Layoffs by Year

We rank the companies with the highest layoffs each year using the DENSE_RANK() function.

```
WITH company_year AS (
   SELECT company, YEAR(`date`), SUM(total_laid_off) AS total_laid_off
   FROM layoffs_staging2
   GROUP BY company, YEAR(`date`)
),
company_year_rank AS (
   SELECT *, DENSE_RANK() OVER(PARTITION BY YEAR(`date`)
ORDER BY total_laid_off DESC) AS Ranking
   FROM company_year
)
SELECT *
FROM company_year_rank
WHERE Ranking <= 5;
```

- **Insight**: This query displays the top 5 companies with the highest layoffs for each year, allowing us to track major players affected by layoffs.

Conclusion:

This Exploratory Data Analysis (EDA) project helps uncover significant insights into global layoffs. We identified the companies with the most layoffs, analyzed industry trends, and explored country-wise impacts. These insights provide a valuable understanding of how layoffs have evolved over time and across different sectors.

By using SQL to perform EDA, we were able to query large datasets efficiently and reveal important data-driven insights. The analysis highlights the impact of economic downturns on companies and industries worldwide, particularly during the challenging years of 2020-2023.

Technologies and Tools Used:

- **SQL**: MySQL for running queries
- **CTEs (Common Table Expressions)**: For modularizing complex queries

- **Window Functions**: For calculating rolling sums and ranks

Repository Link: Exploratory Data Analysis of Layoffs

This document provides a complete step-by-step overview of the analysis performed during the EDA phase of the project.