Layoffs Dataset: Data Cleaning & Exploratory Data Analysis (EDA) Report

Author: Mahmoud

Project: SQL Data Cleaning & EDA

Database: `layoffs`

Tables Used: `layoffs, layoffs_staging, layoffs_staging2`

Tools: MySQL

Data Cleaning Summary

Objective:

Prepare the layoffs dataset for analysis by removing duplicates, standardizing values, handling nulls, and formatting columns.

Cleaning Steps

1. Staging Setup

- Created `layoffs_staging` and `layoffs_staging2` to preserve raw data and track cleaning steps.

2. Duplicate Removal

- Used `ROW_NUMBER()` with `PARTITION BY` to identify exact duplicates.
- Deleted rows with `row_num > 1` to retain unique records.

3. Standardization

- Replaced blank `industry` values with `NULL`.
- Populated missing `industry` using self-join on `company`.
- Normalized inconsistent entries:
- `'CryptoCurrency', 'Crypto Currency'` → `'Crypto'`

- `'United States.'` → `'United States'`

4. Date Formatting

- Converted `date` from text to DATE using `STR_TO_DATE()`.

5. Null Handling

- Retained meaningful nulls in `total_laid_off`, `percentage_laid_off`, and `funds_raised_millions`.
- Removed rows where both `total_laid_off` and `percentage_laid_off` were null.

6. Final Cleanup

- Dropped helper column `row_num`.

Exploratory Data Analysis (EDA)

Objective:

Explore trends, patterns, and outliers in the cleaned layoffs dataset.

Key Insights

1. Largest Layoffs

- Max single-day layoff:
- `SELECT MAX(total_laid_off)`
- Companies with 100% layoffs: Mostly startups (e.g., Quibi, BritishVolt)

2. Funding vs. Failure

- Some companies raised millions yet shut down (e.g., Quibi: $2B)

3. Aggregates

- Top 10 companies by total layoffs

```
GROUP BY company ORDER BY SUM(total_laid_off)
```
- Layoffs by:
- Location
- Country
- Industry
- Stage

4. Temporal Trends

- Yearly layoffs:
```
GROUP BY YEAR(date)
```
- Monthly rolling totals:
```
SUM(total_laid_off) OVER (ORDER BY date)
```

5. Industry Insights

- Most affected industries: Tech, Crypto, Retail
- Standardized industry names for consistency

6. Stage Analysis

- Layoffs by company stage (e.g., Seed, Series A, Public)