JUNE 2025

Global Layoffs Data Cleaning and Exploratory Data Analysis

PRESENTED BY

Babs the Analyst

Table of Content

Global Layoffs Data Cleaning and Exploratory Data Analysis

1. Introduction

In today's data-driven world, organizations rely on accurate and clean data to make decisions. However, raw datasets often come with inconsistencies, duplicates, and missing values that can lead to incorrect analyses. This project aims to clean and explore a layoffs dataset, then extract actionable insights to help understand trends in layoffs across companies, industries, and geographies. It also highlights layoffs in Nigeria, offering local context within the global picture. It demonstrates practical SQL data cleaning techniques and exploratory data analysis (EDA) using advanced SQL concepts such as window functions, CTEs, and aggregation.

2. Project Overview

This project involved two primary stages:

1. **Data Cleaning**: Standardizing and transforming raw data to ensure consistency and completeness.
2. **Exploratory Analysis:** Using SQL queries to reveal trends such as top companies affected, industry impacts, and temporal patterns.

3. Problem Statement

The core problem was that the layoffs dataset contained:

- Duplicate records.
- Inconsistent strings and date formats.
- Missing or blank values.
- Unnecessary columns that hinder analysis.
- This made it difficult to gain accurate insights into the scale and nature of layoffs across different companies, industries, and countries. The problem was to clean and prepare the data using SQL — ensuring consistency, completeness, and reliability for analytical use.

4. Dataset Overview

The dataset contains information on layoffs, including:

- **Company:** Name of the company
- **Location:** City
- **Industry:** Industry sector
- **Total Laid Off**: Number of employees laid off
- **Percentage Laid Off**: Proportion of workforce laid off

- **Date**: Layoff date
- **Stage:** Funding stage of the company
- **Country**: Country where layoffs occurred
- **Funds Raised:** Total funding prior to layoffs

5. Tools Used

- **MySQL:** For data cleaning and SQL queries.
- **SQL Window Functions:** ROW NUMBER () for duplicate removal.
- **Data Formatting Functions:** STR_TO DATE (), TRIM (), GROUP BY, etc.
- **CTEs (Common Table Expressions):** for stepwise aggregation and rolling totals.

6. Data Cleaning Steps

Here's what was done to clean the data:

- **Removed Duplicates**: Used ROW_NUMBER () to identify and delete duplicate rows.
- **Standardized Strings:** Trimmed whitespace and fixed typos in company, industry, country.
- **Formatted Dates**: Converted date to proper SQL DATE format with STR_TO_DATE ().
- **Filled Missing Industries**: Updated blank industry fields by joining on company where information was available.
- **Removed Incomplete Records**: Deleted rows with both total_laid_off and percentage_laid_off as NULL.
- **Dropped Unnecessary Columns:** Removed helper column (row number) after deduplication.

7. Project Structure

Database Setup

The cleaning begins by creating another table layoffs2 with all necessary columns with temporary row_number column is used for deduplication by removing duplicates,

```sql
CREATE TABLE `layoffs2` (
  `company` text,
  `location` text,
  `industry` text,
  `total_laid_off` int DEFAULT NULL,
  `percentage_laid_off` text,
  `date` text,
  `stage` text,
  `country` text,
  `funds_raised_millions` int DEFAULT NULL,
  `row_number` int
) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4 COLLATE=utf8mb4_0900_ai_ci;

insert into layoffs2
 select *,
row_number() over( partition by company, location, industry, total_laid_off, percentage_laid_off,  date, stage,country, funds_raised_millions) as row_num
from layoffs;
```

Data Cleaning Steps

Key cleaning operations in the SQL file:

- **Remove Duplicates**

  Uses the ROW_NUMBER() function to detect and remove duplicates

```sql
delete
from layoffs2
where `row_number` >1;

```

- **Trim Whitespace:** To ensures consistent text values.

```
34 •  select trim(company)
35    from layoffs2;
36
37 •  update layoffs2
38    set company = trim(company);
39
40 •  select trim(location)
41    from layoffs2;
42
43 •  update layoffs2
44    set location = trim(location);
45
46 •  select *
47    from layoffs2;
48
49 •  select trim(industry)
50    from layoffs2;
51
52 •  update layoffs2
53    set industry = trim(industry);
```

```
73
74 •   select country, trim(trailing '.' from country)
75     from layoffs2;
76
77 •   update layoffs2
78     set country = trim(trailing '.' from country)
79     where country = 'states.';
```

```
68 •   select stage, trim(stage)
69     from layoffs2;
70
71 •   update layoffs2
72     set stage = trim(stage);
73
74 •   select country, trim(trailing '.' from country)
75     from layoffs2;
76
77 •   update layoffs2
78     set country = trim(trailing '.' from country)
79     where country = 'states.';
```

- **Convert Data type:** Converts string dates to SQL DATE type

```
55 •   select `date`,
56     str_to_date(`date`, '%m/%d/%Y')
57     FROM layoffs2;
58
59 •   update layoffs2
60     set date = str_to_date(`date`, '%m/%d/%Y');
```

```
55 •   alter table layoffs2
56     modify percentage_laid_off int;
57
```

- **Standardize Values and Fill Missing Data**

```
98      ----  altering of where crypto currency to crypto
99
100 •   update  layoffs2
101     set industry = 'crypto'
102     where industry like 'crypto%';
103
106 •   select * from layoffs2
107     where company = 'airbnb';
108
109 •   UPDATE layoffs2
110      set industry = null
111     where industry = '';
112
113 •   select 1L.industry, 2L.industry
114     from layoffs2 1L
115     join layoffs2 2L
116     on 1L.company = 2L.company
117     where 1L.industry is null
118     and 2L.industry is not null;
119
120 •   update layoffs2 1L
121     join layoffs2 2L
122     on 1L.company = 2L.company
123     SET  1L.industry = 2L.industry
124     WHERE 1L.industry is null
125     and 2L.industry is not null;
126
```

- **Drop Temporary Columns**

```
129 •   alter table layoffs2
130     drop column `row_number`;
131
```

8. Exploratory Data Analysis (EDA)

The second SQL file (layoff query.sql) explores layoff trends using analytical queries. Key EDA components:

a) Summary using DESCRIBE and COUNT () functions.

```sql
--- insightful and exploratory analysis with  layoffs2 dataset using sql
select * from layoffs2;

describe layoffs2;

select count(not null)
from layoffs2;

select count( null)
from layoffs2;

select count(" ")
from layoffs2;
```

b) looking at the start and the end of layoff

```sql
--- looking at the start and the end of layoff
select min(date), max(date)
from layoffs2;
```

c) looking at the start and the end of layoff

```sql
--- looking at the start and the end of layoff
select min(date), max(date)
from layoffs2;
```

9. Insight and Findings

a) Top 5 maximum laid off in one day by company, location, and country.

```sql
select company, location, country,  max(total_laid_off)
from layoffs2
group by company, location, country
order by max(total_laid_off) desc limit 5;
```

b) Top 5 maximum % laid off in one day by company, location, and country.

```
16    --- top 5 maximium % laid off in one day by company, location, and country.
17 •  select company, location, country,  max(percentage_laid_off)
18    from layoffs2
19    group by company, location, country
20    order by  max(percentage_laid_off) desc limit 5;
21
```

c) Total laid off by Company

```
34    ---  total laid off by company
35 •  select company,  sum(total_laid_off)
36    from layoffs2
37    group by company
38    order by  sum(total_laid_off) desc ;
39
```

d) Total laid off by industry

```
40    ---  total laid off by industry
41 •  select company,  sum(total_laid_off)
42    from layoffs2
43    group by industry
44    order by  sum(total_laid_off) desc ;
45
```

e) Which country got the most during layoffs

```
47    --- which country got the most during layoffs
48 •  select country,  sum(total_laid_off)
49    from layoffs2
50    group by country
51    order by  sum(total_laid_off) desc;
52
```

f) Total layoffs by year

```
53      --- total layoffs by year
54 •    select year(date),  sum(total_laid_off)
55      from layoffs2
56      group by year(date)
57      order by  1  desc;
```

g) Total layoffs by month

```
58
59      --- total layoffs by month
60 •    select monthname(date),  month(date),   sum(total_laid_off)
61      from layoffs2
62      where month(date) is not null
63      group by month(date), monthname(date)
64      order by month(date) asc, sum(total_laid_off) desc;
65
```

h) Average % laid off by industry

```
66      --- average % laid off by industry
67 •    select  industry , ceiling( avg(percentage_laid_off))
68      from layoffs2
69      group by industry
70      order by  avg(percentage_laid_off) desc;
71
```

i) which stage got the most during layoffs

```
72      --- which stage got the most during layoffs
73 •    select stage,  sum(total_laid_off)
74      from layoffs2
75      group by stage
76      order by  sum(total_laid_off) desc;
77
```

j) Companies in Nigeria with layoffs

```
78      --- companies in Nigeria with layoffs
79  •   select company, industry, sum(total_laid_off)
80      from layoffs2
81      where country = 'nigeria'
82      group by company, industry
83      order by sum(total_laid_off) desc;
```

k) Total layoff by year in Nigeria

```
85      --- Total layoff by year in Nigeria
86  •   Select YEAR(date) AS year, SUM(total_laid_off) AS total_laid_off
87      FROM layoffs2
88      Where country = 'nigeria'
89      and total_laid_off IS NOT NULL
90      group by YEAR(date)
91      order by year ASC;
```

l) Total layoff by month and year

```
85      --- total layoff by month and year
86  •   select substring(date, 1,7)  month, sum(total_laid_off)
87      from layoffs2
88      where substring(date, 1,7) is not null
89      group by month
90      order by 1 asc;
```

m) Rolling total by month and year

```
--- rolling total by month and year
with rolling_total  as
(select substring(date, 1,7)  month, sum(total_laid_off) laid_off
from layoffs2
where substring(date, 1,7) is not null
group by month
order by 1 asc)
select  month,  laid_off,
sum(laid_off) over (order by month) rolling_total
from rolling_total;
```

10.Final Observations and Insights

Some of the most important findings include:

1. **Tech and consumer companies faced the most severe layoffs**: These sectors consistently topped the list of total layoffs across the dataset. Layoff peaks followed macroeconomic shocks. Sharp spikes in layoffs clearly align with key events like the COVID-19 pandemic in 2020 and the funding downturns in 2022.The steepest layoffs occurred during these turbulent periods, revealing how economic instability propagates into workforce reductions.
2. **Rolling monthly totals highlight sustained patterns of volatility**: By aggregating layoffs into monthly buckets and analyzing the rolling total. The data reveals not just one-off surges, but longer periods of elevated layoffs spanning multiple months.
3. **The rolling total grows rapidly after mid-2020 and mid 2022**: this indicates waves of layoffs spreading across companies rather than isolated incidents. Month-over-month trends help us identify risk windows and anticipate future layoffs if similar economic conditions reoccur.
4. **Seasonality and clustered spikes:** The data also shows seasonal patterns and clustered spikes for example like the End-of-year layoffs often rise as companies tighten budgets. Tech layoffs tend to roll into the first quarter of the next year after investors' funding decisions.
5. **Industries & geographies most affected during spikes**: Industries like tech, consumer internet, and retail saw disproportionate layoffs during peak months. The United States home to most major tech companies also saw the steepest absolute numbers and sharpest spikes in layoffs as compared to other countries.

In Nigeria's Case

- 2022 recorded the highest layoffs (1,020 employees) in Nigeria, aligning with the peak in Jumia's and Alerzo's job cuts within the retail sector.
- The Retail industry dominates layoffs, accounting for over 1,300 cumulative job losses (Jumia and Alerzo) reflecting cost-cutting measures and market competition in e-commerce.
- The Finance sector (Renmoney, Kuda, OPay) follows, with more than 400 layoffs, suggesting restructuring among digital banking and fintech startups.
- The Transportation industry (Gokada, Bolt) faced downsizing as ride hailing operations adjusted post-pandemic.
- Crypto related firms (Nestcoin, Quidax) were also impacted, likely due to volatility in global digital asset markets.
- Smaller firms such as Vendease (Food) contributed modestly to the totals, indicating layoffs spread across multiple emerging industries.

11. Recommendations

Based on the cleaned data and analysis:

1. **Companies and Investors:** Should monitor industry-specific economic signals and diversify their funding sources to mitigate mass layoffs
2. **Employees**: May want to focus on industries with stable or growing demand
3. **Analysts and HR departments:** Implement regular data cleaning and tracking to proactively identify risky trends before they worsen
4. **Policy Makers:** Should explore safety nets for heavily impacted industries and geographies.
5. **Macroeconomic Intervention**: Address inflation, unstable exchange rates, and high logistics costs. Government should strengthen SME financing and ensure foreign exchange stability to help companies manage operational costs without job cuts.

12. Conclusion

This project demonstrated the importance of data cleaning as a critical first step before analysis. The cleaned layoffs data yielded accurate, actionable insights into global layoff patterns, helping stakeholders make better-informed decisions. Continued use of these practices will ensure organizations can respond to workforce changes with greater foresight and resilience.