**Startup Layoffs Dataset (2020 – 2023)**

**Data‑Cleaning & Analysis Documentation**

1. Understand the data. Identify what each column means.

* Company: name of company where layoff occurred
* Location: location of the company
* Industry: type of industry of the company
* Total\_Laid\_Off: number of employees laidoff
* Percentage\_Laid\_Off: percentage of employees laidoff
* Date: Input date of record
* Stage: Indicates the company’s funding or lifecycle stage at the time of the layoff event.

2. Back-up original table.

* 1. Copy original table Layoffs.csv and renamed it as LayoffsCleaned.xlsx.
  2. Converted to .xlsx so that Excel could re‑format dates before MySQL import.

1. Cleaning in Excel.
   1. Remove empty / irrelevant columns.
   2. Fill or standardize values.
      * Dates: converted from MM/DD/YYYY ➜ YYYY‑MM‑DD.
      * Text columns: applied =PROPER(TRIM(cell)) to *Company, Location,* and *Industry*.
   3. **Missing values.**
      * Industry & Stage – looked up other rows with the same *Company*; if still blank, set to Unknown (e.g. *Bally’s Interactive* sourced from the web).
      * Total\_Laid\_Off, Percentage\_Laid\_Off, Funds\_Raised\_Millions – left as NULL; no reliable source.
      * Formula used for imputation: =IF(target="",IFERROR(INDEX(lookup\_range,MATCH(1,(company\_range=company)\*(lookup\_range<>""),0)), "Unknown"), target)
   4. Remove Duplicates.
      * Built Concat Field (=CONCAT(...)) and **Row\_Num** (=COUNTIF(range, value)).
      * Deleted rows with Row\_Num > 1 (duplicates: 400, 411, 956, 2280, 2380).
2. **Import into MySQL.**

CREATE DATABASE IF NOT EXISTS DataCleaningProjects;

USE DataCleaningProjects;

DROP TABLE IF EXISTS Layoffs;

CREATE TABLE Layoffs (

Company VARCHAR(100),

Location VARCHAR(100),

Industry VARCHAR(100),

Total\_Laid\_Off INT,

Percentage\_Laid\_Off DOUBLE,

`Date` DATE,

Stage VARCHAR(100),

Country VARCHAR(100),

Funds\_Raised\_Millions INT

);

-- Verify secure import directory

SHOW VARIABLES LIKE 'secure\_file\_priv';

LOAD DATA INFILE 'C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/LayoffsCleaned.csv'

INTO TABLE Layoffs

FIELDS TERMINATED BY ',' ENCLOSED BY '"'

LINES TERMINATED BY '\n'

IGNORE 1 ROWS

(Company, Location, Industry,

@Total\_Laid\_Off, @Percentage\_Laid\_Off, @`Date`, Stage, Country, @Funds\_Raised\_Millions)

SET Total\_Laid\_Off = NULLIF(TRIM(REPLACE(@Total\_Laid\_Off, '\r','')), ''),

Percentage\_Laid\_Off = NULLIF(TRIM(REPLACE(@Percentage\_Laid\_Off,'\r','')), ''),

`Date` = STR\_TO\_DATE(NULLIF(TRIM(@`Date`), ''), '%c/%e/%Y'),

Funds\_Raised\_Millions = NULLIF(TRIM(REPLACE(@Funds\_Raised\_Millions,'\r','')), '');

-- Verify import

SELECT COUNT(\*) FROM Layoffs;

SELECT \* FROM Layoffs LIMIT 5;

1. **5  Key‑Performance Queries & Findings**
   1. **Overall Impact**

* select count(\*) as Cases, sum(Total\_Laid\_Off) as Employees\_Laid\_Off, round(avg(Percentage\_Laid\_Off),3) as AVG\_PCT\_Laid\_Off from layoffs where Total\_Laid\_Off is not null;
* Result: Cases 1620, Employees 386,079, Average Percentage Layoff 0.224
  1. **YTD Running Total**
     + select date\_format(`Date`, '%Y-%m') as `Month`, sum(Total\_Laid\_Off) as Laid\_Off, sum(sum(Total\_Laid\_Off)) over (partition by year(`Date`) order by date\_format(`Date`, '%Y-%m')) as YTD from layoffs group by year(`Date`), date\_format(`Date`, '%Y-%m') order by date\_format(`Date`, '%Y-%m');
     + Result: From March 2020 to March 2023, roughly 386,000 employees were laid off. The largest wave occurred in January 2023, when more than 84,000 people lost their jobs in a single month. On average, about 10,000 employees were laid off each month, although the numbers varied—some months were quiet, while others spiked, particularly in late 2022 and early 2023.
  2. **Industries Hit Hardest**
* select Industry, sum(Total\_Laid\_Off) as Laid\_Off, count(\*) as Cases, round(avg(Percentage\_Laid\_Off),3) as AVG\_PCT\_Laid\_Off from Layoffs group by Industry order by laid\_off desc;
* Result: The most affected industry was Consumer, which saw 46,782 employees laid off across 119 separate events, with an average layoff rate of 26.3 %. Across all fifteen industries listed, a total of 386,079 employees were let go, and the overall average percentage of workforce laid off was 25.1 %.
  1. **Top 10 Largest Single Cases**
     + select Company, Location, Industry, `Date`, Total\_Laid\_Off, Percentage\_Laid\_Off from layoffs order by Total\_Laid\_Off desc limit 10;
     + Result: The largest single layoff event was recorded at Google in Mountain View, under the Consumer industry. It involved 12,000 employees being laid off on 2023-01-20, with a reported layoff rate of 6.0%. The rest of the top 10 events also involved several thousand employees each, highlighting the significant impact of these reductions across major companies and industries.
  2. **Countries with Most Layoffs**
     + select Country, sum(Total\_Laid\_Off) as Laid\_Off, count(\*) as Cases from layoffs group by Country order by Laid\_Off desc;
     + Result: The United States recorded by far the largest layoffs, with 256,474 employees let go across 1,541 events—around 66 % of all cuts captured in the dataset. In total, the query spans 45 countries and registers 386,079 laid‑off employees from 2,340 separate events. After the U.S., the most affected countries were India (~35,993 layoffs), Netherlands (~17,220), and Sweden (~11,264); each of them represents under 10 % of the overall job losses.