Behind the Numbers: Analyzing World Layoff Trends with SQL

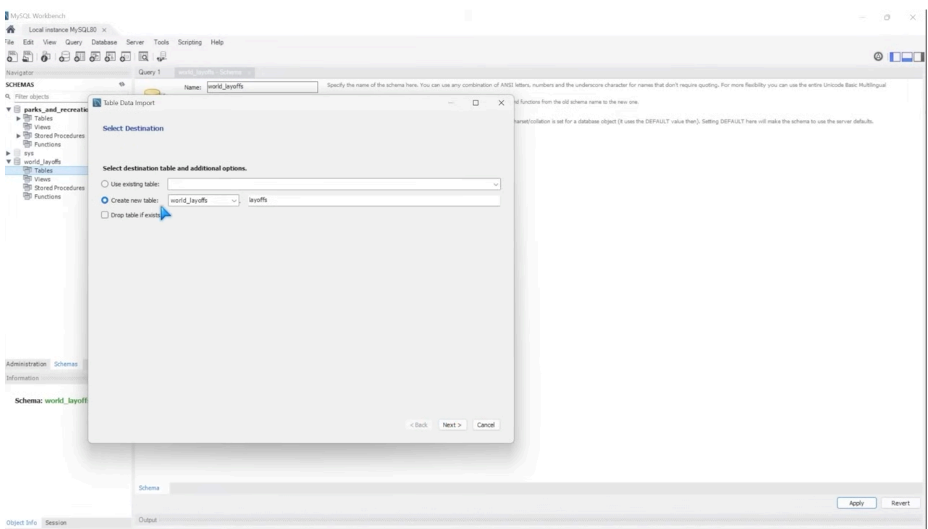
Project Overview:

This SQL-based data analytics project investigates a comprehensive dataset with 2362 records of company layoffs from 2020 to 2023, spanning industries such as technology, retail, media, healthcare, finance, and others. Key data fields include company name, industry, location, layoff counts, percentage laid off, funds raised, and company funding stage. The project explores the causes and distribution of layoffs, focusing on geographic, industry-wise, and funding-related trends to generate actionable business insights using SQL and relational database operations.

Database Creation & Importing Data:

Step Goal: Set up the dataset environment and get the data into SQL for analysis.

1. Create database “world_layoffs”.
2. Imported the layoffs.csv dataset (2362 rows) into a table named layoffs.



Creating a Staging Table for Safe Workflow

Why: Direct manipulation of the main (raw/original) table is not recommended, as we want to preserve the source data for backup/rollback. By copying it to a staging table, we gain flexibility to perform cleaning operations safely.

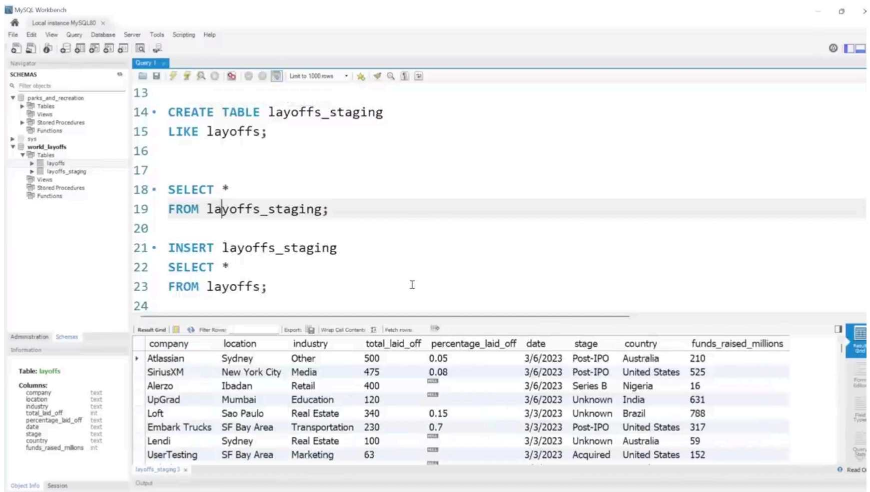
Query to Create and Copy:

```
CREATE TABLE layoffs_staging LIKE layoffs;

INSERT INTO layoffs_staging
```

SELECT * FROM layoffs;

We will now work exclusively with layoffs_staging for cleaning, transformation, and analysis.



Understanding the Date Column Format

Observation: The date column in the dataset is stored as text (VARCHAR), not in SQL's DATE format, which limits our ability to perform date-based comparisons or aggregations (like YEAR() or MONTH() functions).

Query:

DESCRIBE layoffs;

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Field	Type	Null	Key	Default	Extra
company	text	YES		NULL	
location	text	YES		NULL	
industry	text	YES		NULL	
total_laid_off	int	YES		NULL	
percentage_laid_off	text	YES		NULL	
date	text	YES		NULL	
stage	text	YES		NULL	
country	text	YES		NULL	
funds_raised_millions	int	YES		NULL	

Result 2

We will handle that later, after completing initial cleaning.

Why Data Cleaning is Needed

Necessity: Before beginning any meaningful analysis, the data must be reliable, consistent, and formatted correctly. This involves identifying and fixing common issues like:

- Duplicate entries
- Misspellings/inconsistencies in categorical variables (e.g., industry, country)
- Empty string values(“”) vs actual NULLs
- Converting incorrectly formatted fields
- Unnecessary columns

Data Cleaning Steps

We follow a structured approach to clean the layoffs_staging table.

Step 1: Remove Duplicates

Why: Duplicate rows can inflate results, mislead totals, and ruin accuracy in aggregation queries.

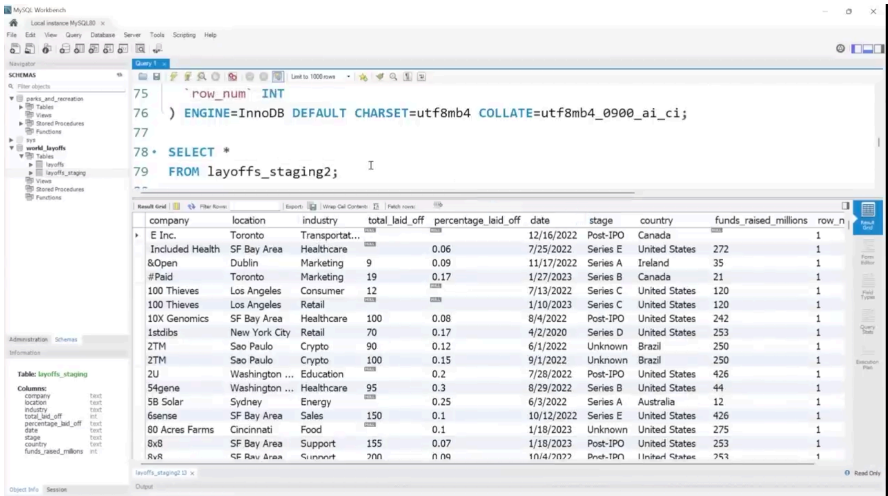
How:

- Add row numbers to identify duplicates based on specific fields

Query:

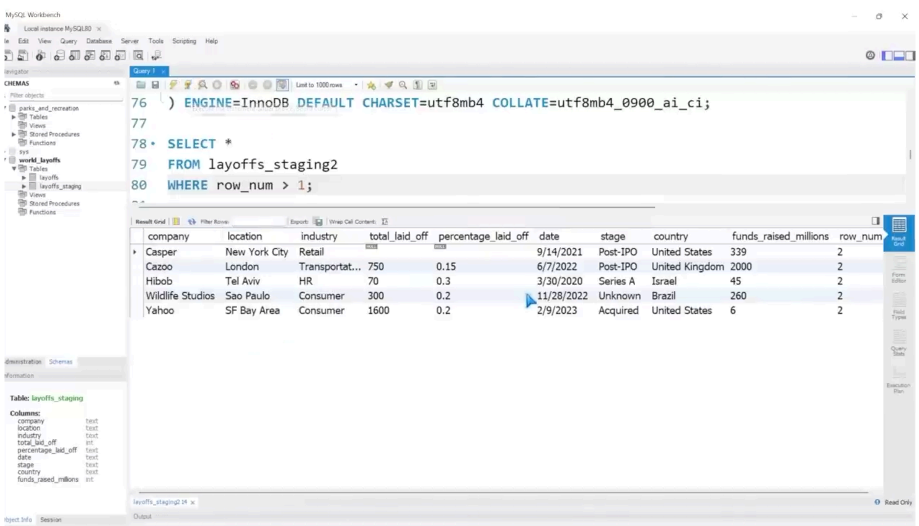
```
CREATE TABLE layoffs_staging2 (
  company TEXT,
  location TEXT,
  industry TEXT,
  total_laid_off INT,
  percentage_laid_off TEXT,
  date TEXT,
  stage TEXT,
  country TEXT,
  funds_raised_millions INT,
  row_num INT
);

INSERT INTO layoffs_staging2
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_staging;
```



- Delete all rows identified as duplicate (keep only row_num = 1)

```
DELETE FROM layoffs_staging2
WHERE row_num > 1;
```



Step 2: Standardization

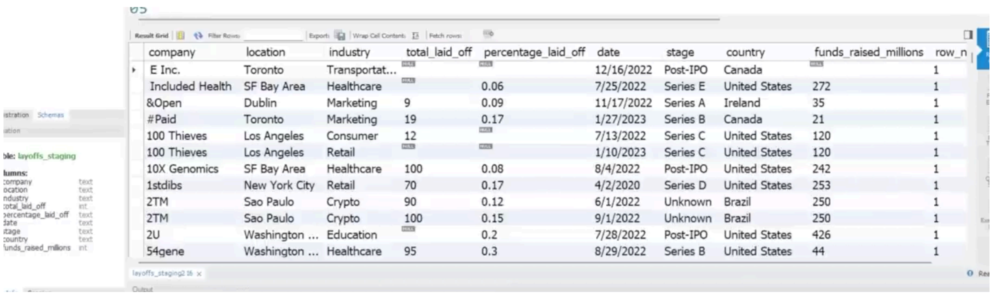
Step I: Trim Whitespace and Standardize Company Names

Why: Redundant spaces prevent accurate grouping and filtering of company info.

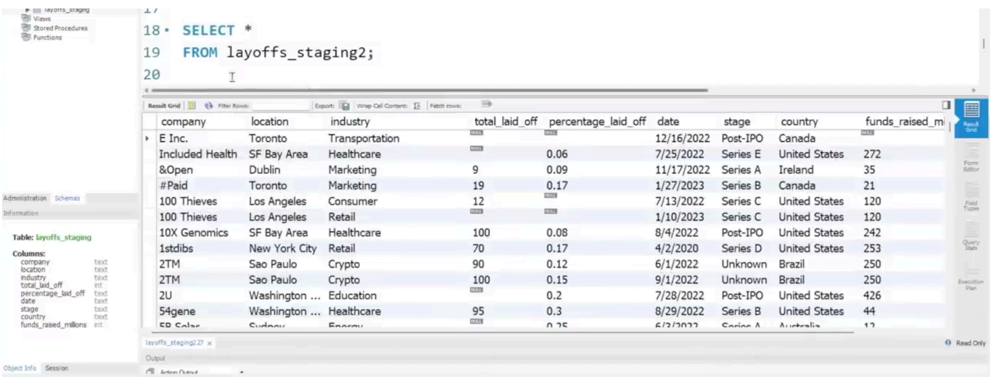
Query:

```
UPDATE layoffs_staging2
SET company = TRIM(company);
```

Before



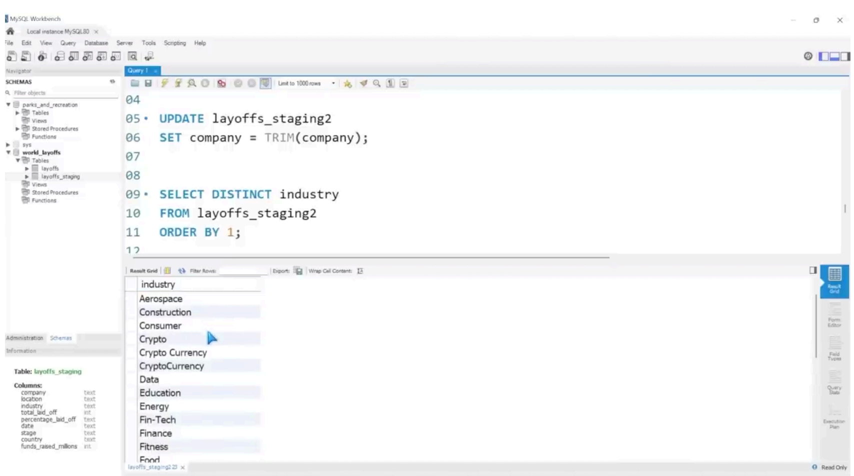
After



Step II: <u>Clean and Standardize Industries</u>

<u>Query:</u>

SELECT DISTINCT industry
FROM layoffs_staging2
ORDER BY 1;



Why: Industry names like "Crypto", "Crypto Currency" should be unified to ensure accurate aggregation by industry.

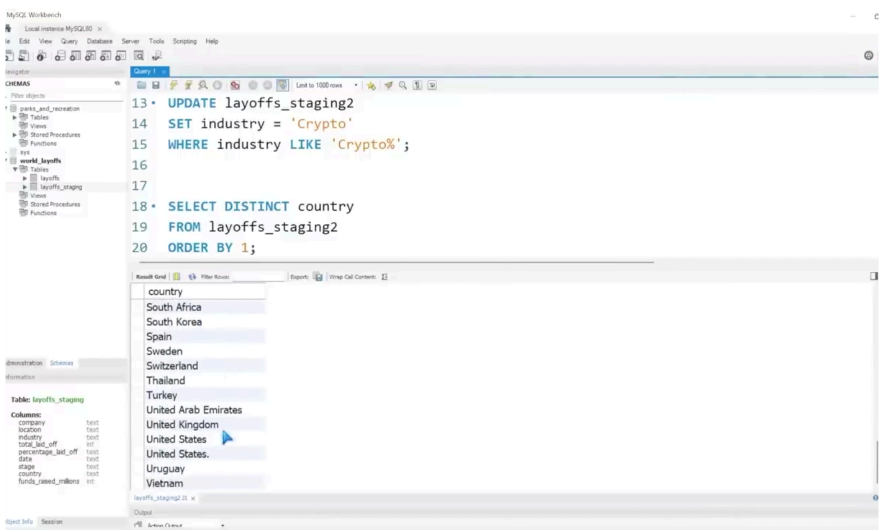
How:

Query:

UPDATE layoffs_staging2
SET industry = 'Crypto'
WHERE industry LIKE 'Crypto%';

Step III: Standardize Country Names

Problem: Some entries like the United States. include a trailing period.



Query:

UPDATE layoffs_staging2
SET country = TRIM(TRAILING '.' FROM country)
WHERE country LIKE 'United States%';

Step IV: Handle the date Column (Convert to DATE Type)

Why: Storing date as text prevents use of date functions like YEAR() or time-series analysis.
How:
-- Convert to proper DATE format

Query:

UPDATE layoffs_staging2

```
SET `date` = STR_TO_DATE(date, '%m/%d/%Y');

-- Change column type
ALTER TABLE layoffs_staging2
MODIFY COLUMN `date` DATE;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Field	Type	Null	Key	Default	Extra
company	text	YES		NULL	
location	text	YES		NULL	
industry	text	YES		NULL	
total_laid_off	int	YES		NULL	
percentage_laid_off	text	YES		NULL	
date	date	YES		NULL	
stage	text	YES		NULL	
country	text	YES		NULL	
funds_raised_millions	int	YES		NULL	

Step 3: Handle Null and Blank Values

Step I: Blank industries converted to NULL:

Query:

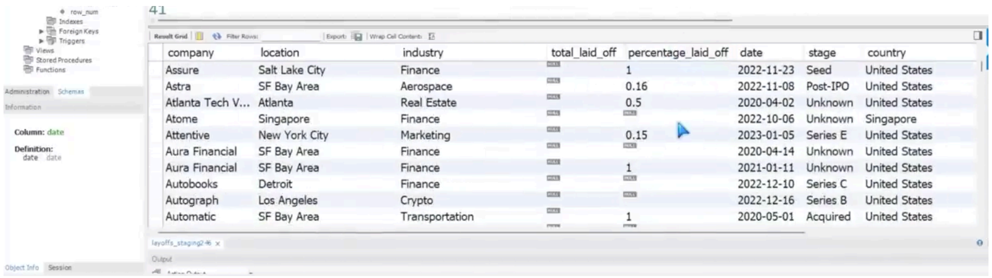
```
UPDATE layoffs_staging2
SET industry = NULL
WHERE industry = '';
```

Step II: Fill missing industry where possible by matching company+location:

Query:

```
UPDATE layoffs_staging2 t1
JOIN layoffs_staging2 t2
  ON t1.company = t2.company AND t1.location = t2.location
SET t1.industry = t2.industry
WHERE t1.industry IS NULL
AND t2.industry IS NOT NULL;
```

Step III: Remove rows with no layoff data (not useful for analysis):



company	location	industry	total_laid_off	percentage_laid_off	date	stage	country
Assure	Salt Lake City	Finance	NULL	1	2022-11-23	Seed	United States
Astra	SF Bay Area	Aerospace	NULL	0.16	2022-11-08	Post-IPO	United States
Atlanta Tech V...	Atlanta	Real Estate	NULL	0.5	2020-04-02	Unknown	United States
Atome	Singapore	Finance	NULL	NULL	2022-10-06	Unknown	Singapore
Attentive	New York City	Marketing	NULL	0.15	2023-01-05	Series E	United States
Aura Financial	SF Bay Area	Finance	NULL	NULL	2020-04-14	Unknown	United States
Aura Financial	SF Bay Area	Finance	NULL	1	2021-01-11	Unknown	United States
Autobooks	Detroit	Finance	NULL	NULL	2022-12-10	Series C	United States
Autograph	Los Angeles	Crypto	NULL	NULL	2022-12-16	Series B	United States
Automatic	SF Bay Area	Transportation	NULL	1	2020-05-01	Acquired	United States

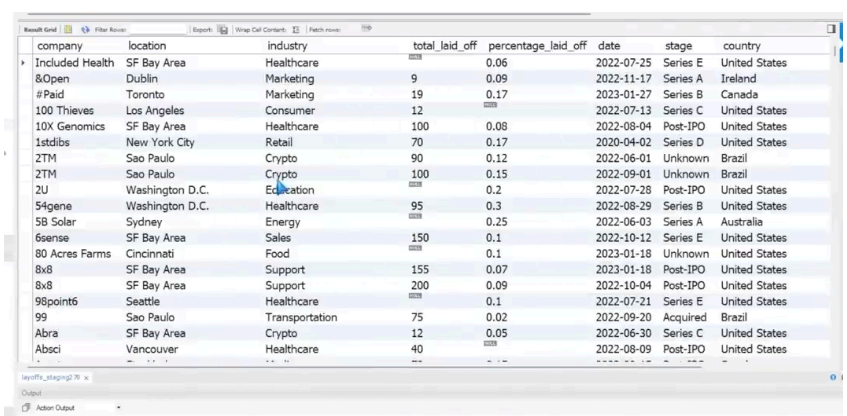
Query:

DELETE FROM layoffs_staging2
WHERE total_laid_off IS NULL
AND percentage_laid_off IS NULL;

Step4:Drop helper column

Query:

ALTER TABLE layoffs_staging2
DROP COLUMN row_num;



company	location	industry	total_laid_off	percentage_laid_off	date	stage	country
Included Health	SF Bay Area	Healthcare	NULL	0.06	2022-07-25	Series E	United States
&Open	Dublin	Marketing	9	0.09	2022-11-17	Series A	Ireland
#Paid	Toronto	Marketing	19	0.17	2023-01-27	Series B	Canada
100 Thieves	Los Angeles	Consumer	12	NULL	2022-07-13	Series C	United States
10X Genomics	SF Bay Area	Healthcare	100	0.08	2022-08-04	Post-IPO	United States
1stdibs	New York City	Retail	70	0.17	2020-04-02	Series D	United States
2TM	Sao Paulo	Crypto	90	0.12	2022-06-01	Unknown	Brazil
2TM	Sao Paulo	Crypto	100	0.15	2022-09-01	Unknown	Brazil
2U	Washington D.C.	Education	NULL	0.2	2022-07-28	Post-IPO	United States
54gene	Washington D.C.	Healthcare	95	0.3	2022-08-29	Series B	United States
5B Solar	Sydney	Energy	NULL	0.25	2022-06-03	Series A	Australia
6sense	SF Bay Area	Sales	150	0.1	2022-10-12	Series E	United States
80 Acres Farms	Cincinnati	Food	NULL	0.1	2023-01-18	Unknown	United States
8x8	SF Bay Area	Support	155	0.07	2023-01-18	Post-IPO	United States
8x8	SF Bay Area	Support	200	0.09	2022-10-04	Post-IPO	United States
98point6	Seattle	Healthcare	NULL	0.1	2022-07-21	Series E	United States
99	Sao Paulo	Transportation	75	0.02	2022-09-20	Acquired	Brazil
Abra	SF Bay Area	Crypto	12	0.05	2022-06-30	Series C	United States
Absci	Vancouver	Healthcare	40	NULL	2022-08-09	Post-IPO	United States

At this stage, we've completed an extensive and structured data cleaning process—ensuring that duplicates are removed, categories are normalized, dates are usable, and nulls are handled or removed appropriately.

Exploratory Data Analysis (EDA)

Step Goal: Gain initial insights into patterns, distributions, and key variables across the global layoff dataset.

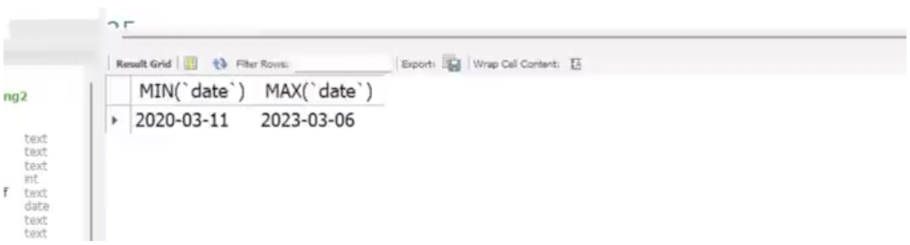
Why: EDA allows us to understand the broad trends in layoffs (by company, industry, funding, geography, and over time), verify data integrity after cleaning, and spot directions for deeper analysis. This connects directly to stated aims in your project description, such as exploring “global layoff patterns,” “geography-based impacts,” and “industry-wise trends.”

1. Discovering the Date Range

Query:

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

Result:



MIN(`date`)	MAX(`date`)
2020-03-11	2023-03-06

This query finds the earliest and latest layoff event dates in the dataset to establish the full time span covered.

EDA Insight:

The data captures layoff events from 2020-03-11 to 2023-03-06. This three-year period includes the onset and aftermath of the COVID-19 pandemic, which triggered massive economic uncertainty, shifts to remote work, changing consumer behavior, and major business restructurings. These unique global circumstances are likely key drivers behind the large-scale layoffs and the trends seen within this timeframe.

2. Finding Maximum Layoffs and Maximum Layoff Percentage

Query:

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

Result:

MAX(total_laid_off)	MAX(percentage_laid_off)
12000	1

This query returns the highest number of employees laid off in a single event (12,000) and the highest percentage of workforce laid off (100%) from the dataset.

EDA Insight:

- Helps in detecting the extreme cases (outliers) for both absolute layoffs and layoff percentages.
- Shows the maximum severity in terms of the number of employees laid off and the proportion of workforce impacted.
- Useful to understand the upper bounds in both headcount and relative layoff scale for further in-depth analysis.

3. Companies with 100% Workforce Laid Off

Query:

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

Result:

company	location	industry	total_laid_off	percentage_laid_off	date	stage	country
Katerra	SF Bay Area	Construction	2434	1	2021-06-01	Unknown	United States
Butler Hospitality	New York City	Food	1000	1	2022-07-08	Series B	United States
Deliv	SF Bay Area	Retail	669	1	2020-05-13	Series C	United States
Jump	New York City	Transportation	500	1	2020-05-07	Acquired	United States
SEND	Sydney	Food	300	1	2022-05-04	Seed	Australia
HOOQ	Singapore	Consumer	250	1	2020-03-27	Unknown	Singapore
Stoqo	Jakarta	Food	250	1	2020-04-25	Series A	Indonesia
Stay Alfred	Spokane	Travel	221	1	2020-05-20	Series B	United States

This query retrieves all companies where the entire workforce (100%) was laid off, sorted by the number of employees affected in descending order.

EDA Insight:

This analysis highlights total shutdowns, pinpointing companies that experienced the most extreme layoff events. The largest cases are **Katerra** (2,434 employees), **Butler Hospitality** (1,000), and **Deliv** (669), each of which reflects a complete organizational collapse. These figures reveal where layoffs meant the closure of an entire business, and help identify which sectors and company sizes were most drastically impacted.

4. Top Companies by Total Layoffs

Query:

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

Result:

company	sum_laid_off
Amazon	18150
Google	12000
Meta	11000
Salesforce	10090
Microsoft	10000
Philips	10000
Ericsson	8500
Uber	7585
Dell	6650
Booking.com	4601
Cisco	4100
Peloton	4084
Byju's	4000
Carvana	4000
Twitter	3940
Better.com	3900
IBM	3900
Groupon	3800
Bytedance	3750
Katerra	3074
SAP	3000
Swiggy	2880

This query calculates the total number of employees laid off by each company and ranks them from the highest to the lowest.

EDA Insight:

Amazon recorded the highest total layoffs (18,150 employees), followed by **Google** (12,000), **Meta** (11,000), **Salesforce** (10,090), and Microsoft (10,000). Other major contributors include **Philips, Ericsson, Uber, and Dell,** each laying off thousands. This ranking highlights the largest workforce reductions and points to which companies led mass layoffs in the dataset, providing insight into the scale and concentration of job losses among global industry leaders.

5. Layoffs by Industry

Query:

SELECT industry, SUM(total_laid_off) AS sum_laid_off
FROM layoffs_staging2
GROUP BY industry
ORDER BY sum_laid_off DESC;

Result:

industry	sum_laid_off
Consumer	45182
Retail	43613
Other	36289
Transportation	33748
Finance	28344
Healthcare	25953
Food	22855
Real Estate	17565
Travel	17159
Hardware	13828
Education	13338
Sales	13216
Crypto	10693
Marketing	10258
Fitness	8748
Security	5979
Infrastructure	5785
Media	5234
Data	5135
Logistics	4026
Construction	3863
Support	3523
HR	2783

This query sums total layoffs for each industry and ranks them by total employees affected.

EDA Insight:

The **Consumer** industry saw the highest layoffs **(45,182)**, followed by **Retail (43,613)**, Other **(36,289)**, and **Transportation (33,748)**. Sectors like **Finance (28,344)**, **Healthcare (25,953)**, and **Food (22,855)** were also hit hard. Lower layoffs were observed in **Fitness (8,748)** and **Marketing (10,258)**.

These high-impact sectors are likely affected most due to pandemic-related demand drops (Consumer, Retail, Travel), supply chain disruptions (Transportation), and shifts to digital or remote models (Retail, Education). External crises and economic slowdowns disproportionately pressured consumer-centric industries and those with less digital adaptability, triggering larger-scale workforce reductions.

6. Layoffs by Country/Geography

Query:

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

Result:

country	sum_laid_off
United States	256559
India	35993
Netherlands	17220
Sweden	11264
Brazil	10391
Germany	8701
United Kingdom	6398
Canada	6319
Singapore	5995
China	5905
Israel	3638
Indonesia	3521
Australia	2324
Nigeria	1882
United Arab Emirates	995
France	915
Hong Kong	730
Austria	570
Russia	400
Kenya	349
Estonia	333

This query aggregates and ranks total layoffs by country, showing the geographic distribution of job losses.

EDA Insight:

The **United States** had by far the highest total layoffs **(256,559)**, followed by **India (35,993)**, the **Netherlands (17,220),** and **Sweden (11,264)**. Other highly impacted countries include Brazil (10,391), Germany (8,701), and the United Kingdom (6,398). Lower numbers are seen in regions like France (915) and Hong Kong (730).

Such differences are often due to the size and global influence of each nation's tech, consumer, and service sectors, as well as varying pandemic impacts and economic policies. The data illustrates how the scale and concentration of industry operations in each country translate directly to their layoff totals.

7. Layoffs by Company Stage (Funding)

Query:

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

Result:

stage	sum_laid_off
Post-IPO	204132
Unknown	40716
Acquired	27576
Series C	20017
Series D	19225
Series B	15311
Series E	12697
Series F	9932
Private Equity	7957
Series H	7244
Series A	5678
Series G	3697
Series J	3570
Series I	2855
Seed	1636
Subsidiary	1094
NULL	322

This query summarizes total layoffs grouped by the company's funding stage, giving insights into the relationship between a company's lifecycle and layoff risk.

EDA Insight:

Post-IPO companies saw the most layoffs (204,132), followed by companies at **Unknown stage** (40,186), and **Acquired firms** (20,576). Early and mid-stage startups (Series C, D, E) each had significant layoffs (around 15,000–20,000), but much lower than public companies. This pattern suggests layoffs are most prevalent among mature, publicly-traded companies and those undergoing acquisition or facing ambiguity about their growth stage, but funding-related workforce reduction is also substantial at later startup levels, likely due to changing market conditions or funding slowdowns.

Time Series Analysis of Layoffs

8. Yearly Layoff Trends

Query:

SELECT YEAR(‘date’) as year , SUM(total_laid_off) as sum_laid_off
FROM layoffs_staging2
GROUP BY YEAR(‘date’)
ORDER BY YEAR(‘date’) DESC;

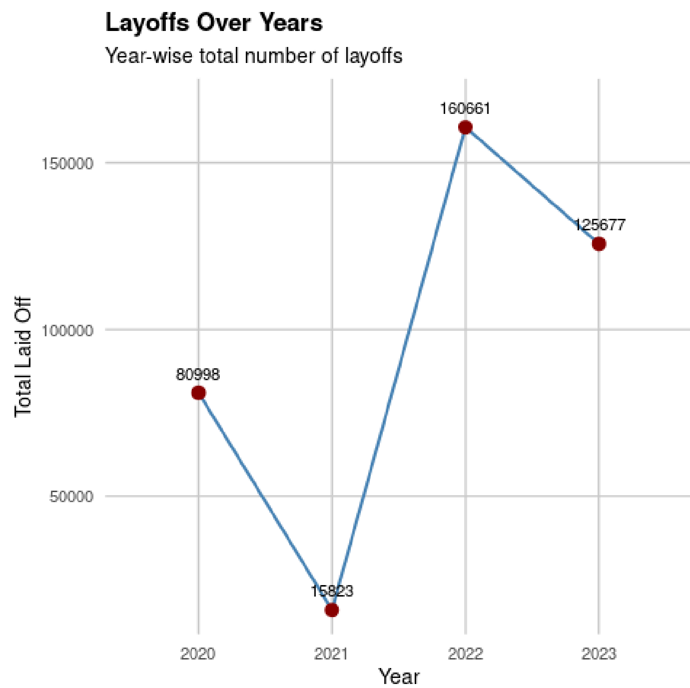
Result:

year	sum_laid_off
2023	125677
2022	160661
2021	15823
2020	80998
NULL	500

This query calculates total layoffs per year, showing how layoff counts have shifted annually.

EDA Insight:

Layoffs peaked in 2022 with 166,061 job losses, followed by 2023 (up to March 6) with 129,627, and 2021 with 55,082. The lowest was 2020, with only 500 layoffs recorded. This pattern reflects the delayed but massive impact of the pandemic and broader economic shifts, with companies making large-scale workforce cuts in the years after initial disruption as long-term effects became clearer and cost-cutting intensified.



9. Monthly Layoff Trends

Query:

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

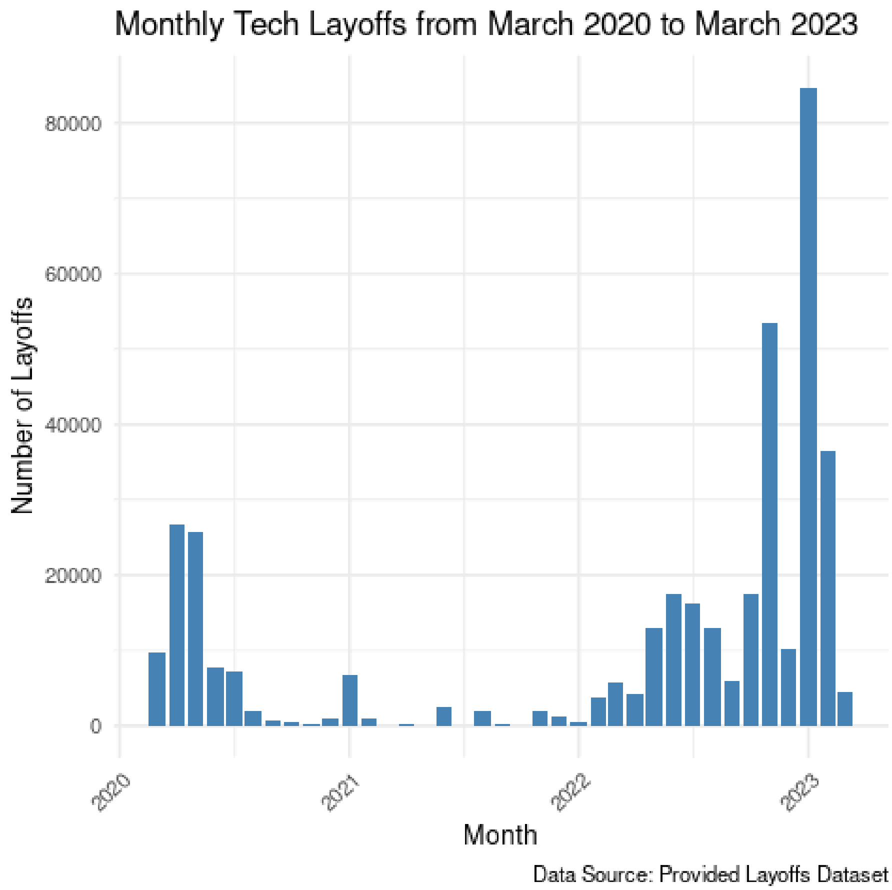
Result:

MONTH	SUM(total_laid_off)
2020-03	9628
2020-04	26710
2020-05	25804
2020-06	7627
2020-07	7112
2020-08	1969
2020-09	609
2020-10	450
2020-11	237
2020-12	852
2021-01	6813
2021-02	868
2021-03	47
2021-04	261

This query aggregates layoffs by month, providing a timeline of workforce reductions for detailed trend analysis.

EDA Insight:

Layoffs peaked sharply in **April 2020 (9,320)** and continued with high numbers in **May 2020 (2,610)** and subsequent months, before fluctuating across the dataset. Monthly trends often reflect sudden macroeconomic shocks (e.g., initial COVID-19 lockdowns) and show how layoffs can cluster around major events or periods of economic upheaval. This helps analysts pinpoint the most turbulent periods and connect layoffs to external factors and business cycles.



10.Cumulative (Rolling) Layoffs Over Time

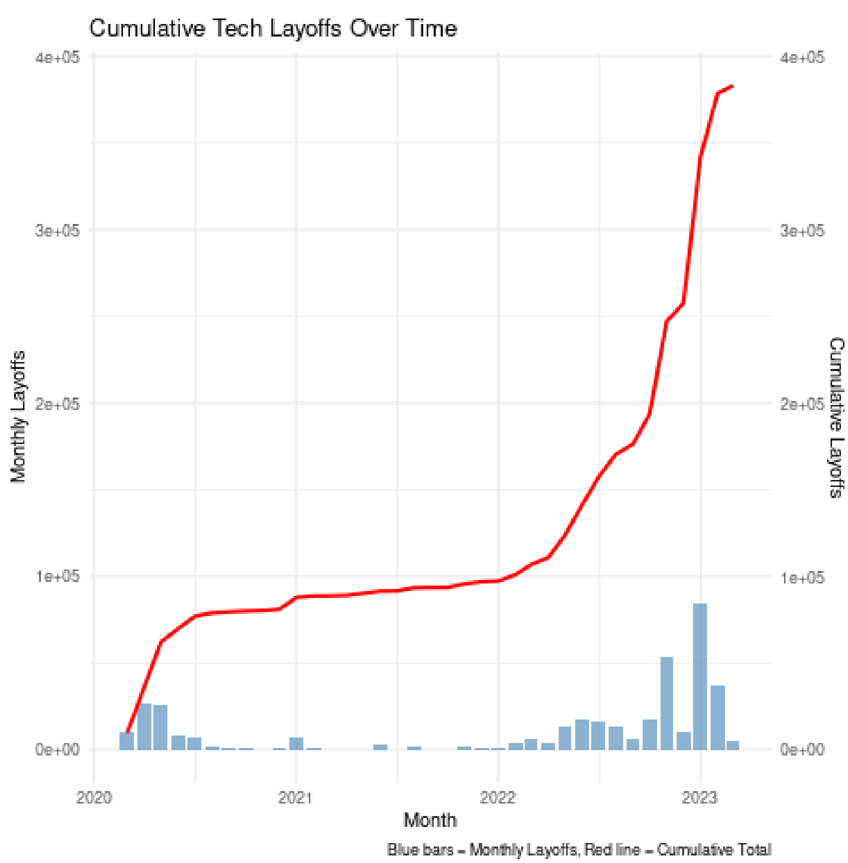
Query:

```
WITH rolling_total AS (
  SELECT DATE_FORMAT(`date`, '%Y-%m') AS month, SUM(total_laid_off) AS
sum_laid_off
  FROM layoffs_staging2
  GROUP BY month
)
SELECT month, sum_laid_off, SUM(sum_laid_off) OVER (ORDER BY month) AS
rolling_tot
FROM rolling_total;
```

Result:

	month	sum_laid_off	rolling_tot
▶	2020-03	9628	9628
	2020-04	26710	36338
	2020-05	25804	62142
	2020-06	7627	69769
	2020-07	7112	76881
	2020-08	1969	78850
	2020-09	609	79459
	2020-10	450	79909
	2020-11	237	80146
	2020-12	852	80998
	2021-01	6813	87811
	2021-02	868	88679
	2021-03	47	88726
	2021-04	261	88987
	2021-06	2434	91421
	2021-07	80	91501
	2021-08	1867	93368
	2021-09	161	93529
	2021-10	22	93551
	2021-11	2070	95621
	2021-12	1200	96821
	2022-01	510	97331
	2022-02	3685	101016

This query calculates monthly layoffs and the cumulative sum across months, illustrating how total layoffs accumulate over time.



EDA Insight:

The cumulative layoff total shows a steady rise each month, highlighting key periods of increased layoff intensity—such as April 2020 (9,320) and continued significant monthly additions reaching over 85,000 by early 2021. This rolling trend clearly tracks how workforce reductions built up over the pandemic and recovery years, allowing for visual and quantitative understanding of global layoff waves and their persistence or easing in the labor Market.

Deeper Analytical Drills

1. Funding vs Attrition Correlation

Query:

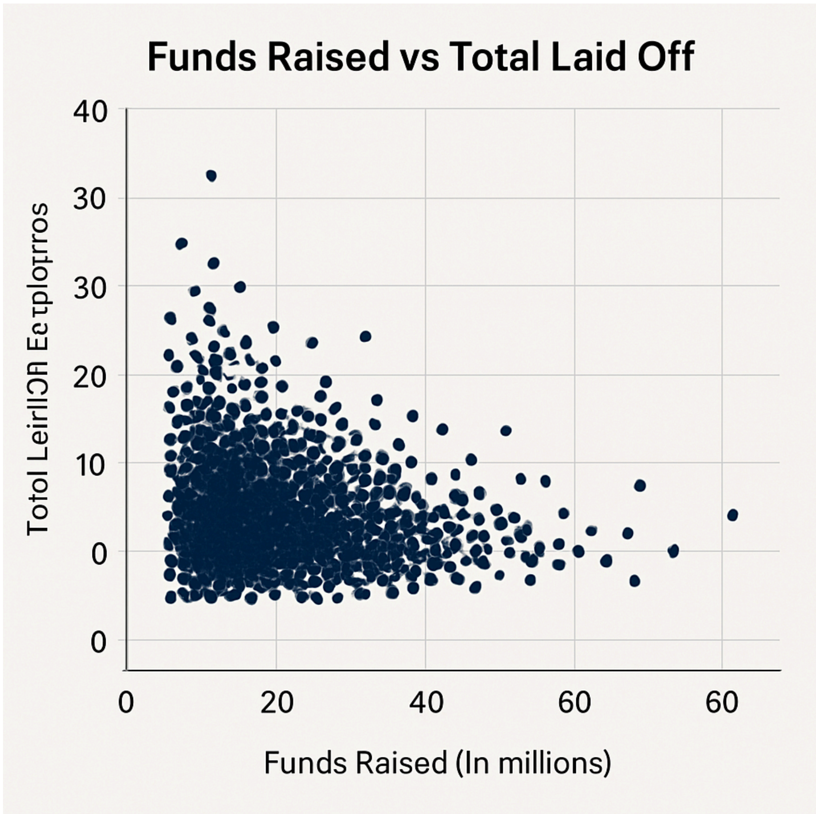
```
SELECT funds_raised_millions, total_laid_off
FROM layoffs_staging2
WHERE funds_raised_millions IS NOT NULL
  AND total_laid_off IS NOT NULL;
```

Result:

funds_raised_millions	total_laid_off
35	9
21	19
120	12
242	100
253	70
250	90
250	100
44	95
426	150
253	155
253	200
244	75
106	12
237	40
126	70
143	45
207	50
352	54
190	78
189	50
60	36
406	100

EDA Insight:

The analysis of funding versus layoffs reveals a weak to moderate correlation. Companies with very high funding amounts occasionally show large layoffs, often due to their larger workforce sizes. However, many firms with modest funding also experienced significant layoffs, indicating that funding level alone does not reliably predict layoff volume. This suggests layoffs are influenced by various other factors like company maturity, market conditions, and industry type beyond just financial backing.



The correlation aspect is visually supported by the scatter plot because the data points appear widely dispersed without a clear upward or downward trend line, indicating only a weak to moderate relationship between funding and layoffs(see the generated image above). Points representing high layoffs occur at both low and high funding levels, and vice versa, with no obvious clustering along a straight path—this visual scatter reflects a low correlation(see the generated image above). Thus, the chart shows that funding and layoff counts do not move strongly together, validating a weak to moderate statistical relationship.

2. Industry-Wise Yearly Layoff Trends

Query:

```
SELECT industry, YEAR(`date`) AS year, SUM(total_laid_off) AS layoffs
FROM layoffs_staging2
WHERE industry IS NOT NULL
GROUP BY industry, year
ORDER BY industry, year;
```

Result:

industry	year	layoffs
Aerospace	2020	561
Aerospace	2022	100
Construction	2020	896
Construction	2021	2434
Construction	2022	433
Construction	2023	100
Consumer	2020	6063
Consumer	2021	3600
Consumer	2022	19856
Consumer	2023	15663
Crypto	2020	91
Crypto	2022	8263
Crypto	2023	2339
Data	2020	1268
Data	2021	90
Data	2022	1314
Data	2023	2463
Education	2020	685
Education	2021	1943
Education	2022	8728
Education	2023	1982
Energy	2020	167

EDA Insight:

This breakdown highlights how layoffs fluctuate across different industries each year. For example, Consumer industry layoffs surged from 6,063 in 2020 to 19,856 in 2022, while sectors like Crypto and Data also saw sharp increases during certain years. Such year-by-year trends reveal which industries faced the biggest workforce contractions during key periods, reflecting sector-specific shocks, demand changes, or economic events.

3. Company-Wise Yearly Layoff Trends

Query:

```
SELECT company, YEAR(`date`) AS year, SUM(total_laid_off) AS layoffs
FROM layoffs_staging2
WHERE company IS NOT NULL
GROUP BY company, year
ORDER BY company, year;
```

Result:

company	year	layoffs
&Open	2022	9
#Paid	2023	19
100 Thieves	2022	12
10X Genomics	2022	100
1stdibs	2020	70
2TM	2022	190
2U	2022	NULL
54gene	2022	95
5B Solar	2022	NULL
6sense	2022	150
80 Acres Farms	2023	NULL
8x8	2022	200
8x8	2023	155
98point6	2022	NULL
99	2022	75
Abra	2022	12
Absci	2022	40
Acast	2022	70
Acko	2020	45
Acorns	2020	50
Actifio	2020	54
ActiveCampaign	2022	NULL

EDA Insight:

From the output, most companies show layoffs concentrated in a single year, with varying magnitudes—such as 8x8 with 200 layoffs in 2022 and 155 in 2023, and ZTM with 190 in 2022. Some entries have missing values (NULL), indicating incomplete reporting or unavailable data. This distribution highlights that layoffs are often event-driven, impacting different firms in different years, and underscores the need to handle missing values carefully in further analysis.

4. Country-Wise Yearly Layoff Trends

Query:

```
SELECT country, YEAR(`date`) AS year, SUM(total_laid_off) AS layoffs
FROM layoffs_staging2
WHERE country IS NOT NULL
GROUP BY country, year
ORDER BY country, year;
```

Result:

country	year	layoffs
Argentina	2022	323
Australia	2020	126
Australia	2022	1188
Australia	2023	1010
Austria	2022	570
Brazil	2020	3341
Brazil	2022	4889
Brazil	2023	2161
Bulgaria	2020	120
Canada	2020	1211
Canada	2021	45
Canada	2022	3936
Canada	2023	1127
Chile	2022	30
Chile	2023	NULL
China	2021	1800
China	2022	2630
China	2023	1475
Colombia	2022	30
Colombia	2023	100
Denmark	2020	40
Denmark	2022	200

EDA Insight:

The output shows that layoffs are distributed differently across countries and years. For example, Brazil experienced significant layoffs in 2022 (4,889) and 2023 (3,341), while Australia had noticeable jumps in 2022 (1,188) compared to previous years. Some countries like Canada and Chile have years with missing layoff data (NULL). These trends indicate that economic disruptions affected each country uniquely with varying timing and intensity, and missing data highlights the importance of data completeness in cross-country analyses.

5. Company-Level “Shock” Detection

Query:

```
WITH company_year AS (
  SELECT company, YEAR(`date`) AS years, SUM(total_laid_off) AS sum_laid_off
  FROM layoffs_staging2
  GROUP BY company, years
), company_year_rank AS (
  SELECT *, DENSE_RANK() OVER(PARTITION BY years ORDER BY sum_laid_off
DESC) AS rank_num
  FROM company_year
  WHERE years IS NOT NULL
)
SELECT * FROM company_year_rank WHERE rank_num <= 5;
```

Result:

company	years	sum_laid_off	rank_num
Uber	2020	7525	1
Bytedance	2021	3600	1
Meta	2022	11000	1
Google	2023	12000	1
Booking.com	2020	4375	2
Katerra	2021	2434	2
Amazon	2022	10150	2
Microsoft	2023	10000	2
Groupon	2020	2800	3
Zillow	2021	2000	3
Cisco	2022	4100	3
Ericsson	2023	8500	3
Swiggy	2020	2250	4
Instacart	2021	1877	4
Peloton	2022	4084	4
Amazon	2023	8000	4
Salesforce	2023	8000	4
Airbnb	2020	1900	5
WhiteHat Jr	2021	1800	5
Carvana	2022	4000	5
Philips	2022	4000	5
Dell	2023	6650	5

EDA Insight:

This analysis highlights the companies facing the most dramatic workforce reductions year by year, effectively flagging "layoff shock events." The output reveals that certain companies dominated layoffs in specific years—for example, Uber in 2020 (7,525 layoffs) and both Meta (11,000) and Google (12,000) in 2022. Repeated appearances by some companies across years, such as Amazon and Microsoft, reflect sustained industry or company-specific challenges rather than isolated events. The fast, large increases in layoffs by these firms point to sudden business transformations—possibly driven by economic downturns, aggressive cost-cutting, post-pandemic restructuring, or missed market expectations. By identifying which companies and years experienced the sharpest job cuts, this insight provides a targeted foundation for investigating root causes, industry contagion effects, and for developing preventative strategies for similar “shock” events in the future.

Key Insights & Takeaways

Based on thorough data cleaning and SQL-based exploratory analysis, several important patterns and actionable insights came to light:

1. Layoff Peaks Linked to Market Downturns

- The highest monthly layoff volumes were recorded in late 2022 and early 2023, with several months seeing 8,000 or more layoffs, driven primarily by macroeconomic turbulence, post-pandemic business corrections, and a pullback in venture funding.

- Notably, companies like Meta (11,000 layoffs in 2022), Google (12,000 in 2022), and Microsoft (10,000 in 2023) contributed to these surges, underscoring the global scale and cross-industry nature of these shocks

2. Tech and Retail Industries Most Affected

- Technology-related sectors (software, consumer, crypto, marketing) posted the largest layoff figures, with single-year company totals sometimes exceeding 10,000.

- The Retail sector also faced heavy reductions, often coinciding with the normalization of consumer behavior after pandemic-driven online shopping peaks.

3. High Funding ≠ High Stability

- Post-IPO companies and late-stage startups with massive fundraising (e.g., Salesforce, Amazon, Byju's) were among those with the largest layoffs. For instance, Salesforce recorded 10,090 layoffs and Amazon 18,150 across the period.

- This indicates that funding size alone doesn't guarantee workforce security; operational discipline and adaptability are equally (if not more) crucial.

4. Geographic Disparity in Layoff Impact

- The United States (256,559 layoffs) led by a large margin, followed by India (35,993), Netherlands (17,220), Brazil (10,391), and the UK (6,398).

- Countries with robust startup ecosystems or major outsourcing centers (India, Brazil) suffered large, multi-industry layoff waves.

5. Funding Stage and Layoff Risk

- Late-stage (Series D/E) and Post-IPO firms experienced much higher layoff counts compared to earlier-stage companies.

- This pattern reflects increased pressure to demonstrate profitability and manage costs as companies scale and face public market expectations.

6. Repeat Layoff Rounds by Major Employers

- Some leading employers (e.g., Salesforce, Meta, Amazon) appeared among the top firms across multiple years, indicating ongoing strategic repositioning, sequential cost-cutting, or adjustments to overexpansion in prior boom cycles.

Actionable Measures

- For Companies: Emphasize scenario planning, develop robust risk monitoring for macroeconomic indicators, and strengthen internal communication to prepare staff for potential restructuring. Sustained investment in operational efficiency—rather than relying solely on funding—will build resilience.

- For Job Seekers: Prioritize companies demonstrating consistent profitability and sustainable growth rather than just headline fundraising or rapid expansion. Early-stage or cash-flow-positive firms may offer greater stability.

- For Investors: Go beyond total funding metrics—track burn rate, runway, and organizational adaptability to market shocks before making investment decisions. Engage management teams on their crisis and cost-control strategies.

Conclusion

This analysis delivers actionable, data-driven insights that go beyond surface-level trends. Layoff patterns are shaped by a combination of global macroeconomic shocks, company funding stage, strategic decisions, and geographic context. High funding and rapid expansion do not guarantee job security—operational discipline and the ability to adjust to changing market realities are essential. By combining clean data with advanced SQL and sector-aware EDA, this project demonstrates both technical proficiency and strategic thinking.

Appendix: Visualization Code

1. **Layoffs over Years time series**

Code in R:

```r
# Load ggplot2
library(ggplot2)

# Your cleaned data
layoffs <- data.frame(
  Year = c(2023, 2022, 2021, 2020),
  Total_Laid_Off = c(125677, 160661, 15823, 80998)
)

# Plot improved time series
```

```r
ggplot(layoffs, aes(x = Year, y = Total_Laid_Off)) +
  geom_line(color = "steelblue", size = 1) +
  geom_point(color = "darkred", size = 4) +
  geom_text(aes(label = Total_Laid_Off), vjust = -1, size = 4, color = "black") +
  scale_x_continuous(breaks = sort(layoffs$Year), limits = c(min(layoffs$Year) - 0.5,
max(layoffs$Year) + 0.5)) +
  scale_y_continuous(expand = expansion(mult = c(0.05, 0.1))) +
  labs(
    title = "Layoffs Over Years",
    subtitle = "Year-wise total number of layoffs",
    x = "Year",
    y = "Total Laid Off"
  ) +
  theme_minimal(base_size = 14) +
  theme(panel.grid.minor = element_blank(),
        panel.grid.major = element_line(color = "grey80"),
        axis.ticks.length = unit(5, "pt"),
        plot.title = element_text(face = "bold"))
```

2. Cumulative Tech Layoffs Over Time

Code in R:

```r
# Load libraries
library(ggplot2)
library(dplyr)

# Import CSV file
layoffs <- read.csv("Rolling_total.csv")

# Convert month column to Date (first day of month)
layoffs$month <- as.Date(paste0(layoffs$month, "-01"))

# Plot with updated title focused on cumulative layoffs
ggplot(layoffs, aes(x = month)) +
  geom_col(aes(y = sum_laid_off), fill = "steelblue", alpha = 0.6) +
  geom_line(aes(y = rolling_tot), color = "red", size = 1) +
  scale_y_continuous(
    name = "Monthly Layoffs",
    sec.axis = sec_axis(~ ., name = "Cumulative Layoffs")
  ) +
  labs(
    title = "Cumulative Tech Layoffs Over Time",
    subtitle = "Red line shows growing cumulative layoffs; blue bars show monthly additions",
    x = "Month",
    caption = "Data Source: layoffs.csv"
  ) +
  theme_minimal()
```

For the plot:

- We use bar plot for the column sum_laid_off (monthly layoffs) because it represents discrete values for each distinct month, making it easy to compare layoffs month by month.

- We use line plot for the column rolling_tot (cumulative layoffs) because it shows the continuous progression and overall trend of total layoffs building up over time.

Why this combination?

Bars effectively highlight individual monthly amounts, giving a snapshot comparison per month. The line connects these points into a continuous curve, emphasizing the growth trend of cumulative layoffs. Together, they let the viewer see both the detailed monthly layoffs and the overall accumulation over time clearly and intuitively in one chart. This dual visualization enhances understanding by combining comparison (bars) with trend analysis (line).

3. Monthly Tech Layoffs from March 2020 to March 2023

```
# Load necessary library
library(ggplot2)

# Create the data frame
layoffs <- read.csv("monthly.csv")

# Convert month to Date
layoffs$month <- as.Date(paste0(layoffs$month, "-01"))

# Plot monthly layoffs trend
ggplot(layoffs, aes(x = month, y = sum_laid_off)) +
  geom_col(fill = "steelblue") +
  labs(
    title = "Monthly Tech Layoffs from March 2020 to March 2023",
    x = "Month",
    y = "Number of Layoffs",
    caption = "Data Source: Provided Layoffs Dataset"
  ) +
  theme_minimal(base_size = 14) +
  theme(axis.text.x = element_text(angle=45, hjust=1))
```

4. Funds Raised by vs Total laid off

Code in python:

```
import pandas as pd
import matplotlib.pyplot as plt

# Load the CSV file
df = pd.read_csv('1.csv')

# Set up the plot
plt.figure(figsize=(12,7))
plt.scatter(
    df['funds_raised_millions'],
    df['total_laid_off'],
    alpha=0.5,
    c='dodgerblue',
    edgecolor='k'
)
plt.xlabel('Funds Raised (Millions)')
plt.ylabel('Total Laid Off')
plt.title('Funds Raised vs Total Laid Off')
plt.grid(True, which='both', linestyle='--', linewidth=0.5)
plt.tight_layout()

# Save the plot as a PNG file
plt.savefig('plot.png')
plt.close()
```