World Layoffs

DATA CLEANING & EDA PROJECT

Feb 2025

Saurav
Payal

Project Overview

Key Points:

- Dataset Source: Kaggle – Global layoffs data during COVID-19.
- Objective: Clean raw data and perform exploratory data analysis (EDA) to uncover trends.
- Technologies Used: MySQL (CTEs, window functions, and advanced SQL techniques).

Data Cleaning Process

```
-- now when we are data cleaning we usually follow a few steps
-- 1. check for duplicates and remove any
-- 2. standardize data and fix errors
-- 3. Look at null values and see what
-- 4. remove any columns and rows that are not necessary - few ways
```

Data Cleaning Process

```sql
SELECT *
FROM (
    SELECT company, industry, total_laid_off,`date`,
        ROW_NUMBER() OVER (
            PARTITION BY company, industry, total_laid_off,`date`
            ) AS row_num
    FROM
        world_layoff.layoffs_staging
) duplicates
WHERE
    row_num > 1;
```

Duplicates- These are the ones we want to delete where the row number is > 1 or 2or greater essentially

company	industry	total_laid_off	date	row_num
Casper	Retail	NULL	9/14/2021	2
Cazoo	Transportation	750	6/7/2022	2
Hibob	HR	70	3/30/2020	2
Oda	Food	70	11/1/2022	2
Oda	Food	70	11/1/2022	3

Exploratory Data Analysis (EDA)

```sql
-- Companies with the biggest single Layoff

SELECT company, total_laid_off
FROM world_layoff.layoffs_staging
ORDER BY 2 DESC;
-- now that's just on a single day
```

company	total_laid_off
Google	12000
Meta	11000
Microsoft	10000
Amazon	10000
Ericsson	8500
Amazon	8000
Salesforce	8000

Google fired the highest number of people in a single layoff.

Exploratory Data Analysis (EDA)

```sql
-- Companies with the most Total Layoffs
SELECT company, SUM(total_laid_off) As total_layoffs
FROM world_layoff.layoffs_staging2
GROUP BY company
ORDER BY 2 DESC
LIMIT 10;
```

company	total_layoffs
Amazon	18150
Google	12000
Meta	11000
Salesforce	10090
Philips	10000
Microsoft	10000
Ericsson	8500

But Amazon fired the most of people!

Exploratory Data Analysis (EDA)

```sql
-- by location
SELECT location, SUM(total_laid_off)
FROM world_layoff.layoffs_staging2
GROUP BY location
ORDER BY 2 DESC
LIMIT 10;

-- this it total in the past 3 years or in the dataset
```

location	SUM(total_laid_off)
SF Bay Area	125631
Seattle	34743
New York City	29364
Bengaluru	21787
Amsterdam	17140
Stockholm	11217
Boston	10785

Most number of people laidoff by City.
(Most of them are US cities)

Exploratory Data Analysis (EDA)

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

country	SUM(total_laid_off)
United States	256559
India	35993
Netherlands	17220
Sweden	11264
Brazil	10391
Germany	8701
United Kinodom	6398

Most number of people laidoff by Country. US by far had the most number of layoffs. (who cares about them anyways :)) (p.s this comment is proof that no one reads these post on linkedin)

Exploratory Data Analysis (EDA)

```sql
SELECT YEAR(date), SUM(total_laid_off)
FROM world_layoff.layoffs_staging2
GROUP BY YEAR(date)
ORDER BY 1 ASC;
```

YEAR(date)	SUM(total_laid_off)
NULL	500
2020	80998
2021	15823
2022	160661
2023	125677

People layed off by Year. As you can see most the numbers reachs its peak in 2022 and then decreases.

Exploratory Data Analysis (EDA)

```sql
WITH Company_Year AS
(
  SELECT company, YEAR(date) AS years, SUM(total_laid_off) AS total_laid_off
  FROM layoffs_staging2
  GROUP BY company, YEAR(date)
)
, Company_Year_Rank AS (
  SELECT company, years, total_laid_off, DENSE_RANK() OVER (PARTITION BY years ORDER BY total_laid_off DESC) AS
  FROM Company_Year
)
SELECT company, years, total_laid_off, ranking
FROM Company_Year_Rank
WHERE ranking <= 3
AND years IS NOT NULL
ORDER BY years ASC, total_laid_off DESC;
```

Layoffs companies wise-year with ranking

Exploratory Data Analysis (EDA)

company	years	total_laid_off	ranking
Uber	2020	7525	1
Booking.com	2020	4375	2
Groupon	2020	2800	3
Bytedance	2021	3600	1
Katerra	2021	2434	2
Zillow	2021	2000	3
Meta	2022	11000	1
Amazon	2022	10150	2
Cisco	2022	4100	3
Google	2023	12000	1
Microsoft	2023	10000	2
Ericsson	2023	8500	3

Layoffs companies wise-year with ranking

Key Insights & Conclusion

Key Points:

- **Major Findings:**
  - **Specific industries and regions were more impacted.**
  - **Identified trends over time that highlight critical periods during the pandemic.**

- **Impact:**
  - **Demonstrated how robust data cleaning drives reliable analysis.**
  - **Enhanced my SQL and data analytical skills.**

Thank you!