Analysis of Global Layoffs - A MySQL Data Cleaning and Exploratory Analysis Project

Executive Summary:

This project involved a detailed analysis of global layoffs using MySQL to manage and analyze large data sets. By leveraging advanced SQL techniques and robust data cleaning methods, the project not only identified key trends and insights into layoff patterns but also showcased sophisticated data management and analytical skills.

company	location	industry	total_laid_off	percentage_laid_off	date	stage	country	funds_raised_millions
Atlassian	Sydney	Other	500	0.05	3/6/2023	Post-IPO	Australia	210
SiriusXM	New York City	Media	475	0.08	3/6/2023	Post-IPO	United States	525
Alerzo	Ibadan	Retail	400	NULL	3/6/2023	Series B	Nigeria	16
UpGrad	Mumbai	Education	120	NULL	3/6/2023	Unknown	India	631
Loft	Sao Paulo	Real Estate	340	0.15	3/3/2023	Unknown	Brazil	788
Embark Trucks	SF Bay Area	Transportation	230	0.7	3/3/2023	Post-IPO	United States	317
Lendi	Sydney	Real Estate	100	NULL	3/3/2023	Unknown	Australia	59
UserTesting	SF Bay Area	Marketing	63	NULL	3/3/2023	Acquired	United States	152
Airbnb	SF Bay Area		30	NULL	3/3/2023	Post-IPO	United States	6400
Accolade	Seattle	Healthcare	NULL	NULL	3/3/2023	Post-IPO	United States	458
Indigo	Boston	Other	NULL	NULL	3/3/2023	Series F	United States.	1200
Zscaler	SF Bay Area	Security	177	0.03	3/2/2023	Post-IPO	United States	148
MasterClass	SF Bay Area	Education	79	NULL	3/2/2023	Series E	United States	461
Ambev Tech	Blumenau	Food	50	NULL	3/2/2023	Acquired	Brazil	NULL
Fittr	Pune	Fitness	30	0.11	3/2/2023	Series A	India	13
CNET	SF Bay Area	Media	12	0.1	3/2/2023	Acquired	United States	20
Flipkart	Bengaluru	Retail	NULL	NULL	3/2/2023	Acquired	India	12900
Kandela	Los Angeles	Consumer	NULL	1	3/2/2023	Acquired	United States	NULL

Introduction:

Background

Global economic fluctuations necessitate a deep understanding of industry and regional layoff trends. Utilizing a comprehensive dataset, this project aims to provide a granular analysis of these trends across various sectors.

Objective

To prepare the dataset for rigorous analysis by meticulously cleaning the data and conducting exploratory data analysis to unearth significant patterns and insights related to industry-specific layoffs and geographical distributions over time.

Data Cleaning Process:

Data Acquisition and Initial Setup

- **Database Setup**: The dataset was imported into the MySQL database world_layoffs using the Table Data Import Wizard, ensuring data integrity from the outset.

Detailed Cleaning Steps

1. **Advanced Duplicate Removal**: Employed window functions to assign unique row numbers for identifying and removing duplicate records, addressing the challenge of lacking unique identifiers and ensuring data accuracy.
2. **Data Standardization and Optimization**:
    - **Standardization**: Implemented TRIM and conditional CASE statements to standardize inconsistent text entries, improving data uniformity across the database.
    - **Type Conversion**: Converted text dates to SQL DATE format using STR_TO_DATE(), enhancing the dataset's usability for time-based analysis.
3. **Null Value Management**: In the project, null values in the industry field were managed using SQL self-joins and update statements, rather than COALESCE(). The technique involved identifying records with null or empty industry fields and filling them by matching against similar records from the same company and location that had complete data. This approach ensured the dataset's completeness and maintained its analytical viability. Follow-up queries were executed to verify that all updates were successfully applied and the data integrity was preserved.
4. **Query Optimization**: Refined SQL queries for performance, including selective indexing to enhance data retrieval times.

Exploratory Data Analysis (EDA):

Analytical Techniques and Methodologies

1. **Maximal and Aggregative Analysis**: Used MAX() to pinpoint extreme layoff cases and grouped data by significant attributes like company and industry using SQL's GROUP BY to identify major layoff contributors.
2. **Time Series and Trend Analysis**:
    - **Window Functions**: Applied SUM() OVER() to compute rolling totals, revealing trends and fluctuations in layoffs over time.
    - **Pattern Recognition**: Analyzed layoffs data to detect industry-specific peaks and troughs, particularly highlighting the impact of COVID-19 on various sectors.

Insights and Implications

- **Industry-Specific Insights**: Detailed analysis revealed that certain industries, notably retail and consumer services, were disproportionately affected by layoffs, highlighting areas for potential economic intervention.
- **Geographical Impact Analysis**: Demonstrated the geographical spread and concentration of layoffs, with the U.S. showing the highest figures, influencing regional economic strategies.

Conclusion:

The project effectively cleansed and analyzed extensive layoffs data, employing advanced SQL functionalities and analytical strategies to derive meaningful insights. These insights not only illuminate past and current trends but also aid in forecasting future industry-specific layoffs.