### Project Report: Data Cleaning and Exploratory Data Analysis (EDA)

#### 1. Data Cleaning

\*\*Initial Extraction and Staging\*\*

- Data was initially extracted from the `layoffsde1..layoffs` table and staged into `layoffs\_staging` to preserve the raw data.

```sql

-- Extracting and staging data

SELECT \*

INTO layoffs\_staging

FROM layoffsde1..layoffs

WHERE 1 = 0;

INSERT INTO layoffs\_staging

SELECT \*

FROM layoffsde1..layoffs;

```

\*\*Duplicate Handling\*\*

- Duplicates were identified based on key attributes (`company, location, industry, total\_laid\_off, percentage\_laid\_off, date, stage, country, funds\_raised\_millions`) and removed to maintain data integrity.

```sql

-- Removing duplicates

WITH duplicate\_cte AS (

SELECT \*,

ROW\_NUMBER() OVER (

PARTITION BY company, location, industry, total\_laid\_off, percentage\_laid\_off, date, stage, country, funds\_raised\_millions

ORDER BY date

) AS CNT

FROM layoffs\_staging

)

DELETE FROM duplicate\_cte

WHERE CNT > 1;

```

\*\*Standardization of Data\*\*

- Company names, industries, locations, and countries were standardized and cleaned to ensure consistency across the dataset.

```sql

-- Standardizing company names

UPDATE layoffs\_staging

SET company = TRIM(company);

-- Standardizing industries

UPDATE layoffs\_staging

SET industry = CASE

WHEN industry LIKE 'Crypto%' THEN 'Crypto'

WHEN industry = 'Transportation' THEN 'Travel'

ELSE industry

END;

-- Standardizing locations

UPDATE layoffs\_staging

SET location = 'Malmo'

WHERE location LIKE 'Malm%';

-- Standardizing countries

UPDATE layoffs\_staging

SET country = TRIM(TRAILING '.' FROM country)

WHERE country LIKE 'United Stat%';

```

\*\*Data Type Conversion and Date Formatting\*\*

- Data types were adjusted where necessary (e.g., converting `percentage\_laid\_off` to FLOAT) and dates were formatted consistently.

```sql

-- Converting percentage\_laid\_off to FLOAT

ALTER TABLE layoffs\_staging

ALTER COLUMN percentage\_laid\_off FLOAT;

-- Formatting dates

ALTER TABLE layoffs\_staging

ALTER COLUMN date DATE;

```

\*\*Handling NULL and Blank Values\*\*

- NULL and blank values in critical fields such as `percentage\_laid\_off` and `total\_laid\_off` were addressed to ensure data completeness.

```sql

-- Handling NULL values in percentage\_laid\_off

UPDATE layoffs\_staging

SET percentage\_laid\_off = CAST(percentage\_laid\_off AS FLOAT)

WHERE ISNUMERIC(percentage\_laid\_off) = 1;

-- Removing rows with NULL values in critical columns

DELETE FROM layoffs\_staging

WHERE percentage\_laid\_off IS NULL AND total\_laid\_off IS NULL;

```

#### 2. Exploratory Data Analysis (EDA)

\*\*Overview of Cleaned Data\*\*

- Initial exploration of the cleaned `layoffs\_staging` dataset to understand its structure and contents.

```sql

SELECT \*

FROM layoffs\_staging;

```

\*\*Analysis of Key Metrics\*\*

- \*\*Maximum Laid Off and Percentage\*\*

- Identifying the highest values for `total\_laid\_off` and `percentage\_laid\_off`.

```sql

SELECT MAX(total\_laid\_off) AS Highest\_Laid\_Off, MAX(percentage\_laid\_off) AS Highest\_Percentage

FROM layoffs\_staging;

```

- \*\*Company-wise Analysis\*\*

- Analyzing total layoffs across different companies to identify major contributors.

```sql

SELECT company, SUM(total\_laid\_off) AS Total\_People\_Laid\_Off

FROM layoffs\_staging

GROUP BY company

ORDER BY Total\_People\_Laid\_Off DESC;

```

- \*\*Industry Impact Analysis\*\*

- Understanding which industries experienced the most layoffs.

```sql

SELECT industry, SUM(total\_laid\_off) AS Total\_People\_Laid\_Off

FROM layoffs\_staging

GROUP BY industry

ORDER BY Total\_People\_Laid\_Off DESC;

```

- \*\*Geographical Analysis\*\*

- Analyzing layoffs by country to identify regions most affected.

```sql

SELECT country, SUM(total\_laid\_off) AS Total\_People\_Laid\_Off

FROM layoffs\_staging

GROUP BY country

ORDER BY Total\_People\_Laid\_Off DESC;

```

- \*\*Temporal Trends\*\*

- Analyzing layoffs over time to identify trends and patterns.

```sql

SELECT YEAR(date) AS Year, SUM(total\_laid\_off) AS Total\_People\_Laid\_Off

FROM layoffs\_staging

GROUP BY YEAR(date)

ORDER BY Year DESC;

```

- \*\*Company Stage Analysis\*\*

- Understanding how layoffs vary across different company stages.

```sql

SELECT stage, SUM(total\_laid\_off) AS Total\_People\_Laid\_Off

FROM layoffs\_staging

GROUP BY stage

ORDER BY Total\_People\_Laid\_Off DESC;

```

- \*\*Rolling Total Analysis\*\*

- Calculating rolling totals of layoffs over time to visualize cumulative impacts.

```sql

WITH Rolling\_Total AS (

SELECT SUBSTRING(date, 1, 7) AS Month, SUM(total\_laid\_off) AS Total\_People\_Laid\_Off

FROM layoffs\_staging

WHERE SUBSTRING(date, 1, 7) IS NOT NULL

GROUP BY SUBSTRING(date, 1, 7)

)

SELECT Month, Total\_People\_Laid\_Off,

SUM(Total\_People\_Laid\_Off) OVER (ORDER BY Month) AS Rolling\_Total

FROM Rolling\_Total;

```

#### Conclusion

The data cleaning phase ensured that the `layoffs\_staging` dataset is accurate, consistent, and ready for in-depth analysis. The exploratory data analysis provided valuable insights into the impact of layoffs across companies, industries, regions, and time periods. These insights are crucial for making informed business decisions, strategic planning, and further data-driven investigations.

Moving forward, advanced analytics techniques such as predictive modeling or deeper sector-specific analyses could provide even deeper insights into the factors influencing layoffs and their broader implications.

This comprehensive approach to data cleaning and exploratory analysis underscores its importance in deriving meaningful insights from raw data, thereby supporting informed decision-making and strategic planning within organizations.