LAYOFFS DATASET

DATA CLEANING

```sql
USE WORLD_LAYOFFS;

SELECT * FROM layoffs;


-- 1. Remove Duplicates

-- 2. Standardize The Data

-- 3. Null Values or Blank Values

-- 4. Remove Any Columns


CREATE TABLE layoffs_staging
LIKE layoffs;


SELECT * FROM layoffs_staging;


INSERT layoffs_staging
SELECT * FROM layoffs;


SELECT * FROM layoffs_staging;
```

CHECK DUPLICATES AND REMOVE IT

```sql
SELECT *,
ROW_NUMBER() OVER(PARTITION BY company, industry, total_laid_off, percentage_laid_off, 'date')
AS row_num
FROM layoffs_staging;


WITH duplicates_cte AS
( SELECT *,
ROW_NUMBER() OVER(PARTITION BY company, location, industry, total_laid_off,
percentage_laid_off, 'date',
stage,country, funds_raised_millions) AS row_num
```

```sql
FROM layoffs_staging)

SELECT * FROM duplicate_cte

where row_num > 1;


SELECT * FROM layoffs_staging

WHERE  company = 'Casper';


WITH duplicates_cte AS

( 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 FROM duplicate_cte

where row_num > 1;


CREATE TABLE `layoffs_staging2` (

 `company` text,

 `location` text,

 `industry` text,

 `total_laid_off` int DEFAULT NULL,

 `percentage_laid_off` text,

 `date` text,

 `stage` text,

 `country` text,

 `funds_raised_millions` int DEFAULT NULL,

 `row_num` INT

) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4 COLLATE=utf8mb4_0900_ai_ci;



SELECT * FROM layoffs_staging2;
```

```sql
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;


SET SQL_SAFE_UPDATES=0;


DELETE FROM layoffs_staging2
WHERE row_num > 1;


SELECT * FROM layoffs_staging2
WHERE row_num > 1;



-- Standardizing Data


SELECT company, TRIM(company)
FROM layoffs_staging2;


UPDATE layoffs_staging2
SET company = TRIM(company);


SELECT DISTINCT industry FROM layoffs_staging2     # Industry Column
WHERE industry LIKE 'Crypto%';                                              # it is same, we need to
change this


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

```sql
SELECT DISTINCT country FROM layoffs_staging2     # Duplicates (Country Column)
ORDER BY 1;


SELECT  DISTINCT country, TRIM(TRAILING '.' FROM country) FROM layoffs_staging2
ORDER BY 1;


UPDATE layoffs_staging2                                          # Duplicates
Remove
SET country = TRIM(TRAILING '.' FROM country)
where COUNTRY like 'United States%';


SELECT `date`,                                                   # Change
date Format
STR_TO_DATE(`DATE`,'%m/%d/%Y')
FROM layoffs_staging2;


UPDATE layoffs_staging2
SET `date` = STR_TO_DATE(`date`, '%m/%d/%Y');          # Update Change Date


SELECT `date`
FROM layoffs_staging2;


ALTER TABLE layoffs_staging2
MODIFY COLUMN `date` DATE;


-- Null Values


SELECT * FROM layoffs_staging2                    # THIS IS NULL VALUES
WHERE total_laid_off IS NULL
AND percentage_laid_off IS NULL;
```

```sql
UPDATE layoffs_staging2
SET industry = null
WHERE industry = '';


SELECT * FROM layoffs_staging2                          # THIS IS NULL VALUES
WHERE industry IS NULL
OR industry = '';


SELECT * FROM layoffs_staging2
WHERE company = 'Airbnb';


SELECT * FROM layoffs_staging2 t1
JOIN layoffs_staging2 t2
ON t1.company = t2.company
where (t1.industry IS NULL OR t1.industry = '')
AND t2.industry IS NOT NULL;


SELECT t1.industry, t2.industry FROM layoffs_staging2 t1                 # fill the blank space into NULL
JOIN layoffs_staging2 t2
        ON t1.company = t2.company
WHERE (t1.industry IS NULL OR t1.industry = '')
AND t2.industry IS NOT NULL;


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

```sql
SELECT * FROM layoffs_staging2                    # COLUMN total_laid_off AND total_laid_off
, all rows are empty

WHERE total_laid_off IS NULL

AND percentage_laid_off IS NULL;


DELETE FROM layoffs_staging2

WHERE total_laid_off IS NULL

AND percentage_laid_off IS NULL;


SELECT * FROM layoffs_staging2;


ALTER TABLE layoffs_staging2

DROP COLUMN row_num;


SELECT * FROM layoffs_staging2;
```