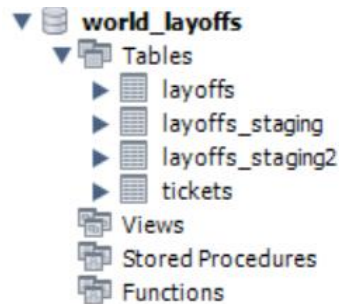


-- Data Cleaning Using MySQL

Worldwide companies layoff data is used



```
SELECT * FROM layoffs;
```

	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

Data Cleaning procedure followed

-- Remove Duplicates

-- Standardize data

-- No values or blank values

-- Remove columns

Create a duplicate table “layoffs_staging”, where the working is performed.

```
CREATE TABLE layoffs_staging
```

```
LIKE layoffs;
```

```
INSERT layoffs_staging
```

```
SELECT * FROM layoffs;
```

```
SELECT * FROM layoffs_staging;
```

```
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;
```

-- Create CTE to find the duplicates in row_num column

```
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) AS row_num  
  
FROM layoffs_staging  
  
)  
  
SELECT *  
  
FROM duplicate_cte  
  
WHERE row_num > 1;
```

```
SELECT *  
  
FROM layoffs_staging  
  
WHERE company = 'Casper';
```

```
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  
WHERE row_num > 1;
```

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

```
SELECT *  
FROM layoffs_staging2;
```

-- Standardizing Data

```
SELECT company, TRIM(company)  
FROM layoffs_staging2;
```

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

```
SELECT DISTINCT(industry)
FROM layoffs_staging2;
```

```
SELECT DISTINCT(industry)
FROM layoffs_staging2
ORDER BY 1;
```

```
SELECT *
FROM layoffs_staging2;
```

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

```
SELECT DISTINCT country
FROM layoffs_staging2
ORDER BY 1;
```

```
UPDATE layoffs_staging2
SET country = 'United States'
WHERE country LIKE 'United States%';
```

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

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

```
SELECT *
FROM layoffs_staging2;
```

```
ALTER TABLE layoffs_staging2
MODIFY COLUMN `date` DATE;
```

-- Remove Null and Blank Values

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

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

```
SELECT *
FROM layoffs_staging2
WHERE industry = null
OR industry = '';
```

```
SELECT *
FROM layoffs_staging2
WHERE Company = 'Airbnb';
```

```
SELECT t1.industry, t2.industry
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;
```

```
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;
```

```
SELECT *
FROM layoffs_staging2;
```

```
SELECT *
FROM layoffs_staging2
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;
```